



# **Spring 2007 Technical Report**



**August 2007**



# Table of Contents

<b>1. Background .....</b>	<b>1</b>
<b>2. Purpose of the Technical Report.....</b>	<b>3</b>
<b>3. Technical Advisory Committee .....</b>	<b>4</b>
<b>4. Transition of Contract.....</b>	<b>6</b>
4.1 Replication of the Spring 2006 Scoring .....	6
4.2 Communication Plan .....	16
4.2.1 Technology Focus Group.....	16
4.2.2 Report Design Focus Group.....	16
4.2.3 ISAT Overview Workshops.....	16
4.2.4 Report Content, Use, and Interpretation Webex Sessions .....	17
4.2.5 Fall/Winter Computer Adaptive Testing Focus Group.....	17
<b>5. New Item Development .....</b>	<b>20</b>
5.1 Multiple-choice Items.....	22
5.2 Sampling of the Content Domain .....	22
5.3 Test Blueprints.....	24
5.4 Item Writing .....	25
5.5 Passage Selection and Readability .....	28
5.6 Universal Design .....	28
5.7 Depth of Knowledge.....	29
5.8 Item Content Review .....	29
5.9 Sensitivity and Bias Review .....	31
5.10 Forms Approval Committee .....	32
5.11 Security.....	32
5.12 Maintenance of the Item Bank.....	33
<b>6. Spring 2007 Embedded Field Test .....</b>	<b>35</b>
6.1 Purpose .....	35
6.2 Distractor Analysis .....	35
6.3 Item Analysis – Differential Item Functioning (DIF) .....	40
6.4 Rasch Item Analysis .....	45

<b>7. Operational Test Forms Construction.....</b>	<b>50</b>
7.1 2007 Operational Plan .....	50
7.2 Forms Construction Tools and Documentation.....	54
<b>8. Test Administration.....</b>	<b>59</b>
8.1 Sessions .....	59
8.2 Accommodations .....	60
8.3 Test Security .....	60
<b>9. Quality Control Processing.....</b>	<b>61</b>
<b>10. Student Examinee Demographic Summary — Census.....</b>	<b>62</b>
<b>11. Operational Item and Form Summary.....</b>	<b>71</b>
11.1 Distribution of P-Values and Item-Total Correlations by Grade .....	71
11.2 Rasch Summaries for the Core Form .....	75
11.3 Raw and Scale Score Descriptive Information for the Core Forms .....	80
11.4 Student Scale Score Distributions Across Grades .....	89
<b>12. Reliability.....</b>	<b>94</b>
12.1 Standard Error of Measurement .....	105
12.2 Indicators of Consistency .....	109
<b>13. Performance Levels .....</b>	<b>111</b>
<b>14. Linking.....</b>	<b>113</b>
<b>15. Scaling.....</b>	<b>124</b>
15.1 Raw to Scale Score Conversions .....	124
<b>16. Validity.....</b>	<b>126</b>
16.1 External Third-Party Content Alignment Study .....	126
16.2 Content Validity .....	126
16.3 Intercorrelations-Convergent/Divergent Validity.....	127
16.4 Relationship Between the Core and Extender Scores .....	141
<b>17. Proficiency Level Descriptors and Standards Validation.....</b>	<b>151</b>
17.1 Proficiency Level Descriptors .....	151
17.2 Standards Validation.....	151
<b>References .....</b>	<b>152</b>

## Appendices

<b>A – Replication Study Data .....</b>	<b>A–1</b>
<b>B – Overview Workshop PowerPoint Presentation .....</b>	<b>B–1</b>
<b>C – Report Content, Use, and Interpretation Webex Training Sessions.....</b>	<b>C–1</b>
<b>D – Focus Group Panelists and List of Questions &amp; Responses.....</b>	<b>D–1</b>
<b>E – Item Review Materials .....</b>	<b>E–1</b>
<b>F – Item Tracking Review Form .....</b>	<b>F–1</b>
<b>G – Fairness in Testing Manual.....</b>	<b>G–1</b>
<b>H – Sensitivity Review Form.....</b>	<b>H–1</b>
<b>I – Forms Approval Tracking Form .....</b>	<b>I–1</b>
<b>J – Confidentiality Agreement .....</b>	<b>J–1</b>
<b>K – Distractor Analysis.....</b>	<b>K–1</b>
<b>L – Formulas for MH-statistic &amp; Categorization Rules.....</b>	<b>L–1</b>
<b>M – Item Level DIF Statistics .....</b>	<b>M–1</b>
<b>N – Rasch Item Parameters &amp; Fit Statistics.....</b>	<b>N–1</b>
<b>O – Spring 2007 ISAT Test Administration Manual.....</b>	<b>O–1</b>
<b>P – Spring 2007 ISAT District Survey Results.....</b>	<b>P–1</b>
<b>Q – Multiple-choice Distractor Analysis.....</b>	<b>Q–1</b>
<b>R – Rasch Item Parameters and Fit Statistics.....</b>	<b>R–1</b>
<b>S – Linking Steps.....</b>	<b>S–1</b>
<b>T – Linking Plots .....</b>	<b>T–1</b>
<b>U – Raw to Scale Conversions.....</b>	<b>U–1</b>
<b>V – High-level Test Blueprints .....</b>	<b>V–1</b>
<b>W – Standards Validation Report .....</b>	<b>W–1</b>
<b>X – Participation Rates .....</b>	<b>X–1</b>
<b>Y – Committee Meeting Details .....</b>	<b>Y–1</b>



# 1. Background

Article IX, Section 2 of the Idaho Constitution vests general supervision of the state educational institutions and the public school system in the Idaho State Board of Education (Board). School districts, including specially chartered school districts are under the supervision and control of the Board (Section 33-116, Idaho Code). The Board is responsible to prescribe minimum curriculum requirements and determine how textbooks and other curricular materials are adopted (Section 33-118, Idaho Code). In addition, the Board is required by Section 33-1612 to adopt rules establishing a thorough system of public schools. These rules are published through the requirements of the Idaho Administrative Procedures Act (IDAPA) process and have the force of law.

IDAPA 08.02.03.111.01 sets out the philosophy for assessment in Idaho.

**Philosophy.** Acquiring the basic skills is essential to realization of full educational, vocational, and personal/social development. Since Idaho schools are responsible for instruction in the basic scholastic skills, the State Board of Education has a vested interest in regularly surveying student skill acquisition as an index of the effectiveness of the educational program. This information can best be secured through objective assessment of student growth. A statewide student assessment program consisting of standardized achievement testing and performance appraisal activities in the fundamental basic skills will be conducted annually. The State Board of Education will provide oversight for all components of the comprehensive assessment program...

The Office of the State Board of Education (OSBE) carries out the administrative duties of the Board.

Consequent to the passage of 2001 reauthorization of the federal Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), in January 2002 the Board of Education entered into a contract with Northwest Evaluation Association (NWEA) to develop and administer the Idaho Standards Achievement Tests (ISAT), which are a component of an overall student testing program in the state. The purpose of this test was to meet the requirements of NCLB. The contract was for two years with three one-year optional extensions available. The first pilot of the ISAT occurred in the spring of 2002, and the first full administration delivered for the purpose of determining adequate yearly progress (AYP) took place in the spring of 2003.

All three options to extend the contract were accepted. During the final year of the contract (2006), in accordance with a recommendation from the Idaho Technical Advisory Committee, plans included publishing a request for proposal early in the year so that a new contract could be awarded by mid-year. This mid-year award was intended to provide an overlap in contracts to assure a smooth transition in the event that the same vendor was not awarded a second contract.

As required by NCLB the Board contracted with a third party in the spring of 2005 to conduct a study of the alignment, reliability, and validity of the ISAT. This work was accomplished by the Human Resources Research Organization (HumRRO). The report of this study indicated that the ISAT had issues in all three areas. The test was not aligned to the structure of the Idaho content standards, items for the test were selected on the basis of item response theory difficulty values without benefit of data that reflected the quality of the items, data finally provided indicated significant problems with items used, normative averages by grade level not aligned with the goal of accountability systems intended to improve student achievement, and the reporting of scores should be aligned to the standards. (See HumRRO, *Idaho Standards Achievement Test: Review of Evidence of Content Validity*, Released May 2005.)

In September of 2005 the assessment system underwent a federal peer review as required by NCLB. In December the Board was notified that the assessment system not only did not meet the requirements of NCLB but also did not meet the requirements of the prior authorization of the ESEA for which Idaho was under a compliance agreement. The state was fined for failure to comply with the requirements of the prior authorization and ordered to move forward to make corrections for compliance with NCLB. The first requirement was to develop a timeline for compliance, and the second requirement was to undergo a second peer review in September 2006. Corrections that needed to be made were to set cut scores by a technically sound methodology, to align tests to standards, to conduct an alignment study of the spring 2007 core forms, revisit the cut scores to validate them based on the new test following the spring 2007 administration, assure year to year comparability of tests and scores.

By June 2006 the Division of Purchasing, Idaho Department of Administration, the independent purchasing arm of the state system, had determined that a new vendor would be awarded the contract for continuing ISAT. NWEA would continue its contract through the end of the year and deliver the fall 2006 ISAT. The new vendor, Data Recognition Corporation (DRC) along with its partner Computerized Assessments and Learning (CAL), began to bring Idaho's test into compliance with nationally recognized standards for large scale assessment and federal requirements. The current contract runs through July 2010 with two possible extensions of two years each.

In July 2006 DRC brought together Idaho educators to set new cut scores by an acceptable methodology. The new cut scores were included in the Board's documentation for the peer review in the fall of 2006. With commitments to complete the tasks begun according to the original timeline, the U.S. Department of Education approved with recommendations Idaho's assessment system in November 2006. When test items were provided by NWEA to DRC for the transition, two facts became clear: 1) All of Idaho's items were jointly owned by NWEA so Idaho would have only temporary use of them; and 2) there were insufficient numbers of test items that were aligned to Idaho standards to complete the 2007 administration of ISAT. Test development activities began in earnest. A field test of items for grades 3-8 and 10 was conducted in December 2006. Since then over 700 Idaho educators have been involved in various aspects of test development: item review; bias and sensitivity review; form development; development of performance level descriptors by content, by grade, by performance level; an alignment study; and achievement standard setting. The first administration of Idaho's revised ISAT took place in the spring of 2007.

This technical manual documents Idaho's development and administration of an assessment that is technically sound and that meets the requirements set for a statewide assessment system.

The ISAT measures academic content standards in reading, mathematics, and language usage for students in grades 2-10 and in science for grades 5, 7, and 10. The grade 10 ISAT has also served as a high school graduation test since the graduating class of 2006. All test items are in a multiple choice format. The reading test includes selected passages with associated items. The ISAT is administered by computer, although special accommodated versions of the test are provided in large print, Braille, and paper-pencil.



## **2. Purpose of the Technical Report**

The purpose of this report is to provide the Idaho OSBE, educators, citizens, researchers, and other interested parties with technical documentation for the development, administration, and reporting of the Spring 2007 ISAT including evidence of the reliability, validity, and the appropriate use and interpretation of test scores.

This technical report specifically covers the Spring 2007 administration of the ISAT only. Future technical reports will be created for future administrations.

### 3. Technical Advisory Committee

The technical advisory committee (TAC) for the Idaho assessment programs convenes in person three times per year. As needed, conference calls may be used to bridge the gap as issues arise between regularly scheduled meetings. Its current members include:

- Dr. Thomas Fisher

Dr. Fisher holds a BS in Mathematics from Middle Tennessee State University, a M.Ed. in Secondary Administration from the University of Toledo, and an Ed.D. in Curriculum Development from Wayne State University in Detroit, Michigan. He was a mathematics teacher for eight years, a school district program evaluation specialist for two years, and the Coordinator of Dissemination and Training for the Michigan Educational Assessment Program for four years. He served 26 years as the Florida Department of Education's Educational Testing and Evaluation Administrator responsible for K-12 student testing programs, college-level testing programs, and professional licensure examination programs.

Dr. Fisher has served as an advisor on assessment and accountability issues to the U.S. Department of Education and several state education agencies. He has published over 60 articles in professional journals and made many presentations at national and regional professional meetings.

- Dr. William Erpenbach

Dr. Erpenbach holds a Ph.D. in counseling and guidance with a minor in educational administration at the University of Wisconsin, Madison. He was a senior administrator with the Wisconsin Department of Public Instruction (WDPI) and has served as an advisor on assessment and accountability issues to the U.S. Department of Education and several state education agencies. Dr. Erpenbach has written extensively regarding critical issues related to NCLB and how States and the U. S. Department of Education (ED) have implemented the law; served as a peer reviewer for the U. S. Department of Education's for numerous examinations of States' standards, assessments, and accountability systems. He currently serves on technical advisory committees for four states and the World-Class Instructional Design and Assessment Consortium.

- Dr. Edward Roeber

Dr. Roeber holds a BA in Psychology from University of Michigan, an MA in Educational Psychology from University of Michigan, and a Ph.D., Measurement and Evaluation from University of Michigan. He is currently professor, Measurement and Quantitative Methods, with an adjunct appointment in the College of Education, Michigan State University. Dr. Roeber has served as executive director for assessment and accountability for the Michigan Department of Education; vice-president for external relations, Measured Progress; and, director, student assessment programs, Council of Chief State School Officers.

- Dr. Joseph Ryan

Dr. Ryan holds an AB in mathematics, M Ed in Educational Psychology from Boston College, and a PhD in Measurement, Evaluation, and Statistical Analysis from the University of Chicago. He has been an assessment advisor and measurement consultant for more than 20 years. He has

worked with schools, school districts, state departments of education, and many test developers and is currently a member of several Technical Advisory Committees. Dr. Ryan's areas of technical expertise include scaling, equating, standard setting, and bias or DIF analyses.

The role of the TAC is to advise and provide recommendations to OSBE on the ongoing technical issues in the Idaho assessment programs. At the request of OSBE, TAC members may also be asked to provide recommendations or author papers as relevant to their expertise or role as advisors. The regularly scheduled meetings include OSBE and state assessment staff, along with vendor staff.

The meeting typically involves presentations of technical information related to the assessment programs by test vendors and an overview of recent and forthcoming Board actions. Issues range from test administration, item and test development, accommodations and special populations, NCLB requirements, state policy, and psychometrics (reliability, validity, linking). Examples of topics from previous meetings related to the ISAT include:

- Review and Evaluation of the Spring ISAT Technical Report
- Review of the Fall Field Test Sample Plan
- Standard Setting, Standards Validation, and Development of the Performance Level Descriptors
- Replication Study Results
- Linking the New ISAT to the Old ISAT
- Goal and Purpose of the Computer Adaptive Testing Extender Session
- Design of the Fall, Winter, and Summer Test Administrations

An agenda is available before each meeting, along with technical documentation and specific questions that will be presented to the TAC. A report of the meeting is compiled and distributed soon after the close of the meeting. Its contents are reviewed for accuracy before it is finalized. As a means of providing continuity across meetings, the report of the meeting from the previous TAC is included in the current TAC materials.

## 4. Transition of the Contract

Upon award of the contract, DRC requested files and data from OSBE and the previous vendor, Northwest Evaluation Association (NWEA). With permission from OSBE, DRC worked directly with NWEA to get needed information quickly and efficiently. This was to ensure appropriate steps would be taken to transfer necessary records and data from the previous vendor to DRC. As part of the transition plan, OSBE leased 671 items from NWEA for use on the ISAT through February 2008.

The following are the activities that surround the transition of a testing program from one vendor to the next. They include 1) the replication of the Spring 2006 scoring and item calibrations as part of the quality control plan to verify the capabilities of the new vendor, and 2) the communication of changes in the program to the stakeholders in Idaho.

Transition activities in this chapter are structured as follows:

- Replication of the Spring 2006 Scoring
  - Study One: Raw to Scale Conversions
  - Study Two: Raw to Scale Conversions
  - Spring 2006 Item Calibrations
- Communication Plan
  - Technology Focus Group
  - Report Design Focus Group
  - ISAT Overview Workshop
  - Report Content, Use and Interpretation Web-ex Sessions
  - Fall/Winter Computer Adaptive Testing Focus Group

### 4.1 Replication of the Spring 2006 Scoring

As part of DRC's contract for the ISAT, a replication study was conducted. Its goal was two-fold: one, to replicate the scoring from the previous spring (2006) using the same data file as the previous vendor and two, to verify the item calibrations of the Idaho-owned items in the ISAT item bank. The former was verification that DRC would be able to apply the same logic and processes in preparation for their Spring 2007 test implementation. The latter was to assist in the selection of the items that would be used for linking the 2007 Spring ISAT to the 2006 Spring ISAT in order to establish a common metric. A residual benefit of the second part of the study was that it assisted in selecting items for the July 2006 standard setting. Based on the results of the latter study, for each grade and content, a minimum of one and a maximum of three items were excluded from the set of standard setting materials.

To complete the required analysis replication study, DRC was provided the following information by the previous contractor:

- Rasch item difficulties for all items in the pre-equated item bank.
- Sparse data set (scored and unscored student responses) with omits and items not administered indicated for both the calibration sample and the total population.
- Linear transformations from the logit to the scale score metric.
- List of rules or conventions for rounding.
- List of inclusion rules for students.
- Final raw to logit to scale score conversion tables.
- Conditional standard errors.
- Final student file with non-scored and scored item responses, calculated raw scores for both total and subscores, and scale scores.

The replication study was conducted to verify that DRC is capable of establishing the baseline scale using the existing Spring 2006 student data set provided by OSBE and in cooperation with the former vendor under their transition agreement. While it was not expected that this effort would result in a complete duplication, the goal was to be able to replicate the 2006 results within an acceptable degree of precision. In addition, in DRC's study, application of the existing linear transformations, along with appropriate rounding rules, was performed to convert the raw scores to scale scores via the logit metric. Raw to scale conversion tables, along with associated conditional standard errors, were also computed and verified. To further ensure the validity of this study, DRC's partner, Computerized Assessment and Learning, LLC (CAL) conducted an independent study using the same method and data set. The results from these studies, named Study One and Study Two, follow.

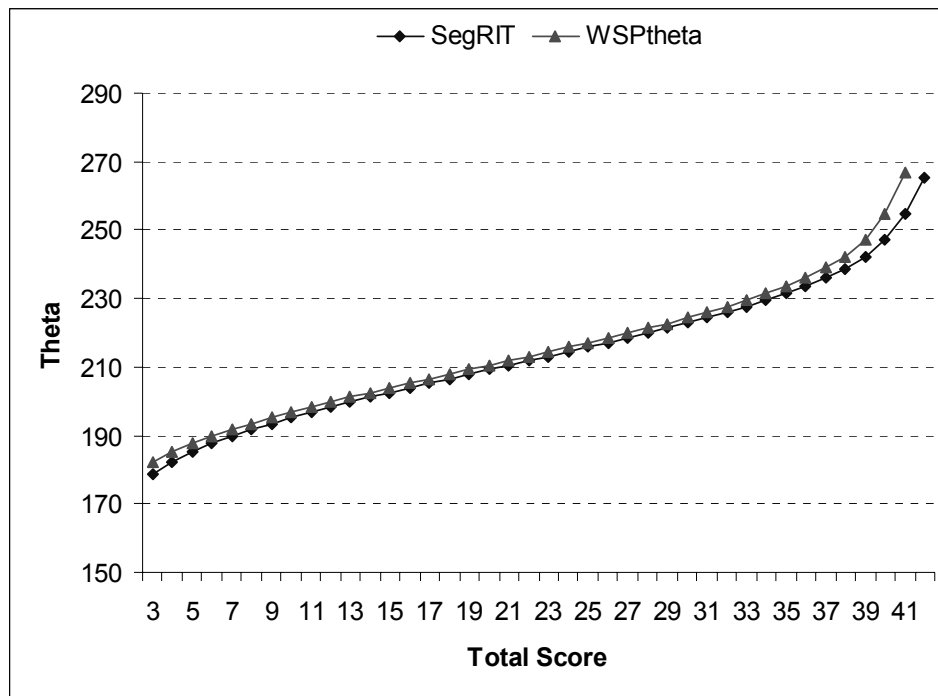
## **Results: Study One**

CAL utilized a direct approach by fixing the scaled item difficulty parameters with the values DRC had been sent from the former vendor and then determining student scale scores from WINSTEPS (Linacre, 2002) using these "fixed" item parameters. The results were systematic across all data files. CAL was able to match the student scale scores (within .03 units on average) for virtually all total score values. The exceptions were for students with perfect (100% correct) scores and students with zero or near-zero scores. In both cases, it is hypothesized that the former vendor used a method to assign these extreme scores outside of WINSTEPS (e.g., hand or analytic smoothing), or by other means (e.g., some portion of a standard error). For the purpose and scope of this study, these extreme scores were not investigated further.

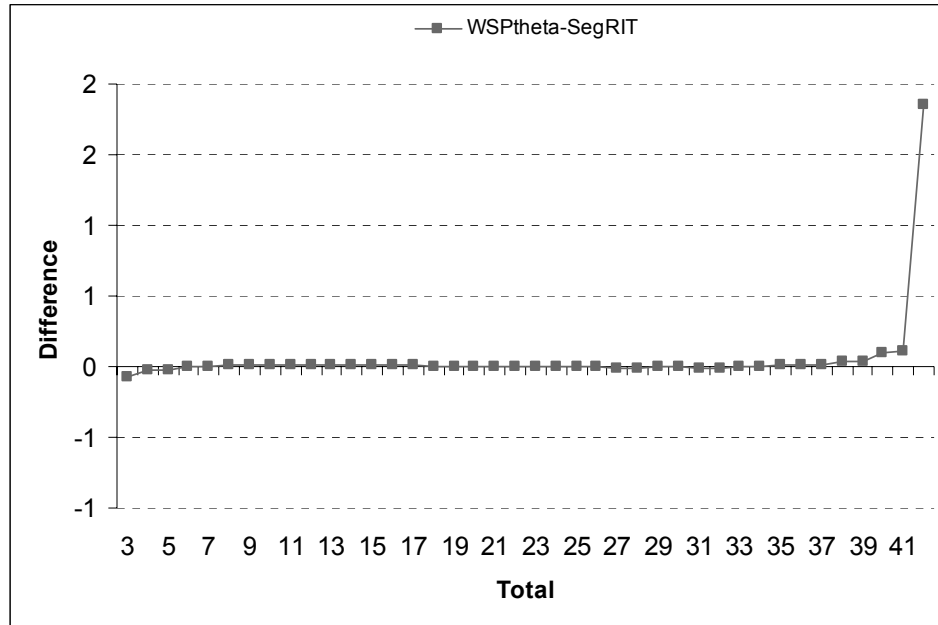
Below is an example of the plots and data results for grade 6 in both reading and mathematics. Appendix A contains the full set of data for all grades in reading and mathematics. SegRIT represents the former vendor values and WSPtheta represents those values determined by CAL. The vertical axis is in the scale score metric and the horizontal axis is in the raw score metric. The differences between the two curves in Figure 4.1 represent the scale score discrepancies. In Figure 4.2, these differences are displayed in the logit metric, with logit discrepancies along the vertical axis and raw score along the horizontal.

Without the presence of systematic bias, the differences should sum to zero across all values and be randomly distributed on either side of zero. Note in the plot that the highest obtainable raw score is noticeably higher for WSPtheta (CAL's estimate). This is consistent across all grades and contents. For grade 6 mathematics (Figure 4.3 and Figure 4.4), the agreement is closer.

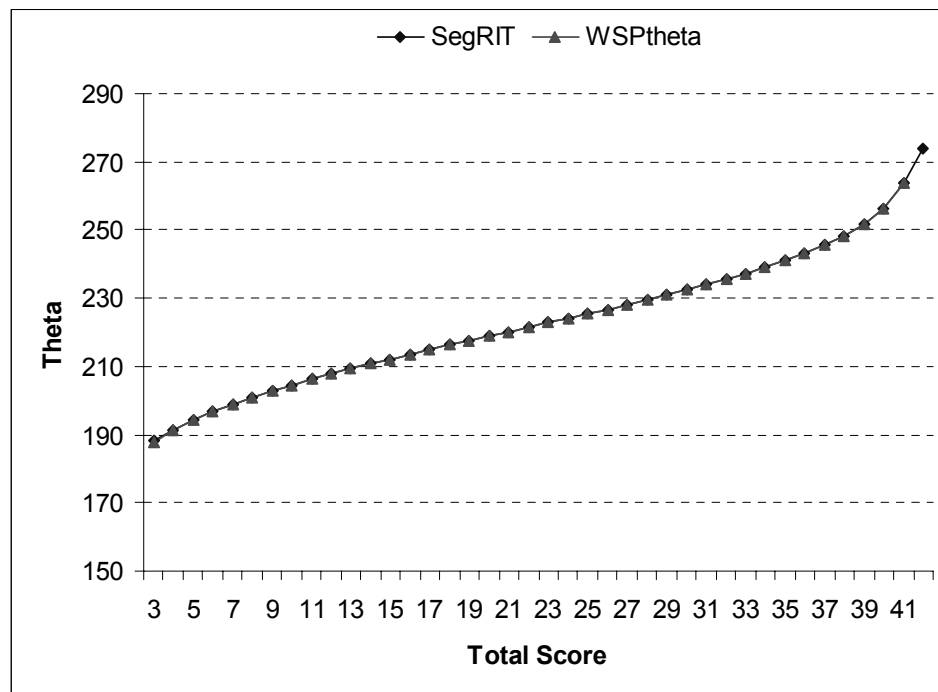
**Figure 4.1. Raw Score versus Scale Score—Original versus Replicated**  
**Grade 6 Reading**



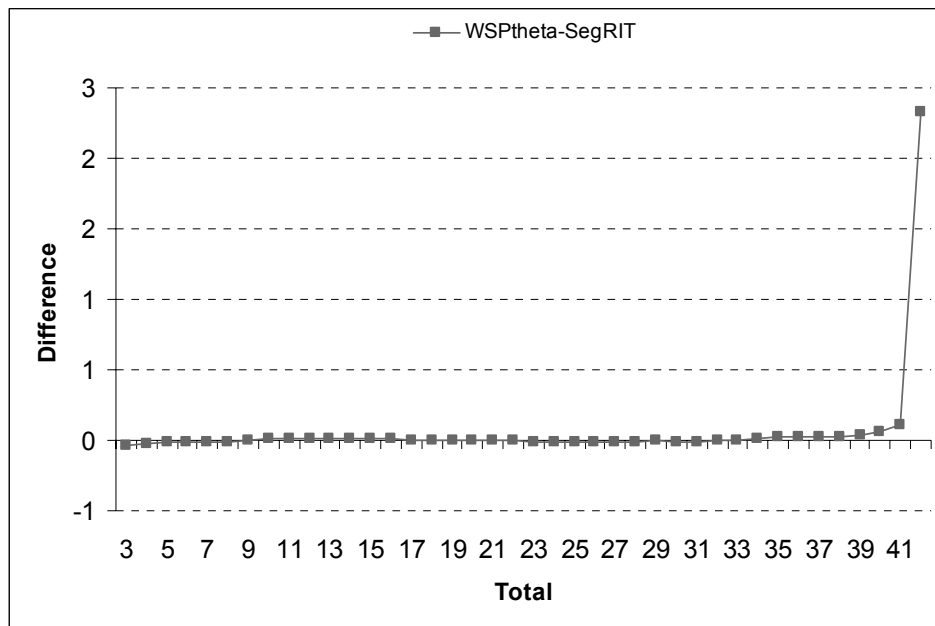
**Figure 4.2. Logit Discrepancies versus Raw Score—Original versus Replicated  
Grade 6 Reading**



**Figure 4.3. Raw Score versus Scale Score—Original versus Replicated  
Grade 6 Mathematics**



**Figure 4.4. Logit Discrepancies versus Raw Score—Original versus Replicated  
Grade 6 Mathematics**



## Results: Study Two

DRC utilized a similar approach to the CAL study. Below is an example of DRC's output for the same grade and contents as above. As in the CAL study, the DRC results matched those of NWEA for all save the extreme scores (note cell highlighted in yellow). For extreme scores, DRC utilized the convention of subtracting .3 of a raw score point from perfect scores and adding .3 of a raw score point to zero scores. This estimation was performed within Winsteps. Table 4.1 shows the results for grade 6 reading and Table 4.2 shows the results for grade 6 mathematics.



**Table 4.1. Raw Score and Scale Score Match between DRC and NWEA**  
**Grade 6 Reading**

DRC			NWEA		Match
Raw	Ability	Scale	Raw	RIT	
0	-4.6	154			N/A
1	-3.3	167			N/A
2	-2.6	174			N/A
3	-2.1	179	3	179	Yes
4	-1.8	182	4	182	Yes
5	-1.5	185	5	185	Yes
6	-1.2	188	6	188	Yes
7	-1.0	190	7	190	Yes
8	-0.8	192	8	192	Yes
9	-0.7	193	9	193	Yes
10	-0.5	195	10	195	Yes
11	-0.3	197	11	197	Yes
12	-0.2	198	12	198	Yes
13	0.0	200	13	200	Yes
14	0.1	201	14	201	Yes
15	0.3	203	15	203	Yes
16	0.4	204	16	204	Yes
17	0.5	205	17	205	Yes
18	0.7	207	18	207	Yes
19	0.8	208	19	208	Yes
20	0.9	209	20	209	Yes
21	1.1	211	21	211	Yes
22	1.2	212	22	212	Yes
23	1.3	213	23	213	Yes
24	1.4	214	24	214	Yes
25	1.6	216	25	216	Yes
26	1.7	217	26	217	Yes
27	1.8	218	27	218	Yes
28	2.0	220	28	220	Yes
29	2.1	221	29	221	Yes
30	2.3	223	30	223	Yes
31	2.4	224	31	224	Yes
32	2.6	226	32	226	Yes
33	2.8	228	33	228	Yes
34	2.9	229	34	229	Yes
35	3.1	231	35	231	Yes
36	3.4	234	36	234	Yes
37	3.6	236	37	236	Yes
38	3.9	239	38	239	Yes
39	4.2	242	39	242	Yes
40	4.7	247	40	247	Yes
41	5.5	255	41	255	Yes
42	6.7	267	42	265	No

**Table 4.2. Raw Score and Scale Score Match between DRC and NWEA**  
**Grade 6 Mathematics**

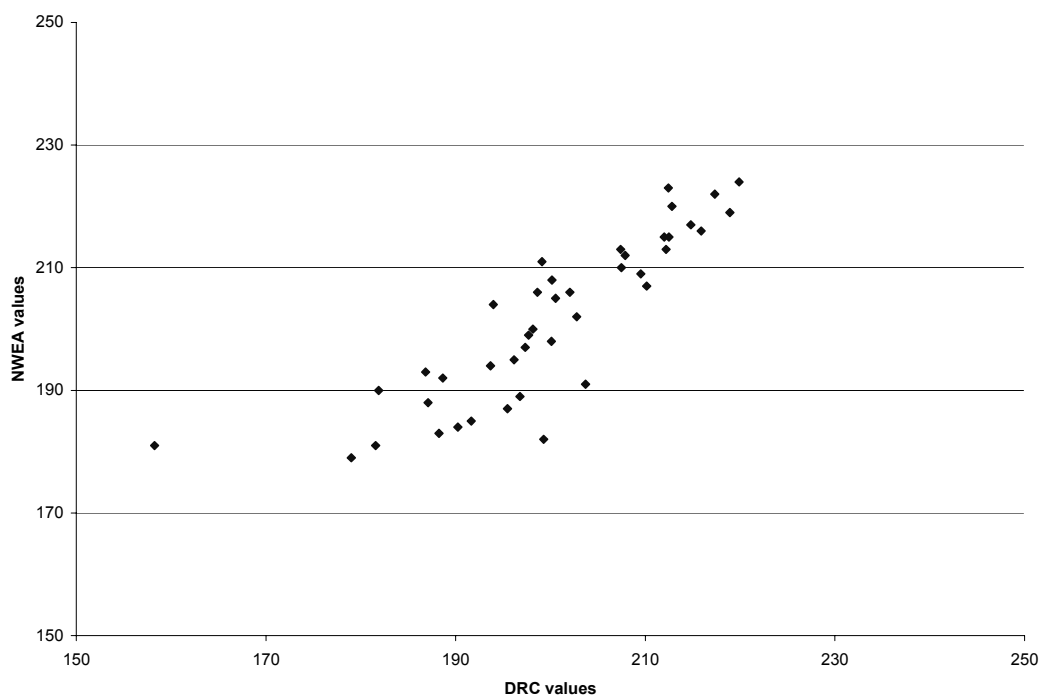
DRC			NWEA		Match
Raw	Ability	Scale	Raw	RIT	
0	-3.7	163			N/A
1	-2.4	176			N/A
2	-1.7	183	2	183	Yes
3	-1.2	188	3	188	Yes
4	-0.8	192	4	192	Yes
5	-0.6	194	5	194	Yes
6	-0.3	197	6	197	Yes
7	-0.1	199	7	199	Yes
8	0.1	201	8	201	Yes
9	0.3	203	9	203	Yes
10	0.5	205	10	205	Yes
11	0.6	206	11	206	Yes
12	0.8	208	12	208	Yes
13	0.9	209	13	209	Yes
14	1.1	211	14	211	Yes
15	1.2	212	15	212	Yes
16	1.4	214	16	214	Yes
17	1.5	215	17	215	Yes
18	1.6	216	18	216	Yes
19	1.8	218	19	218	Yes
20	1.9	219	20	219	Yes
21	2.0	220	21	220	Yes
22	2.1	221	22	221	Yes
23	2.3	223	23	223	Yes
24	2.4	224	24	224	Yes
25	2.5	225	25	225	Yes
26	2.7	227	26	227	Yes
27	2.8	228	27	228	Yes
28	3.0	230	28	230	Yes
29	3.1	231	29	231	Yes
30	3.2	232	30	232	Yes
31	3.4	234	31	234	Yes
32	3.6	236	32	236	Yes
33	3.7	237	33	237	Yes
34	3.9	239	34	239	Yes
35	4.1	241	35	241	Yes
36	4.3	243	36	243	Yes
37	4.6	246	37	246	Yes
38	4.8	248	38	248	Yes
39	5.2	252	39	252	Yes
40	5.6	256	40	256	Yes
41	6.4	264	41	264	Yes
42	7.6	276	42	274	No

## Spring 2006 Item Calibrations

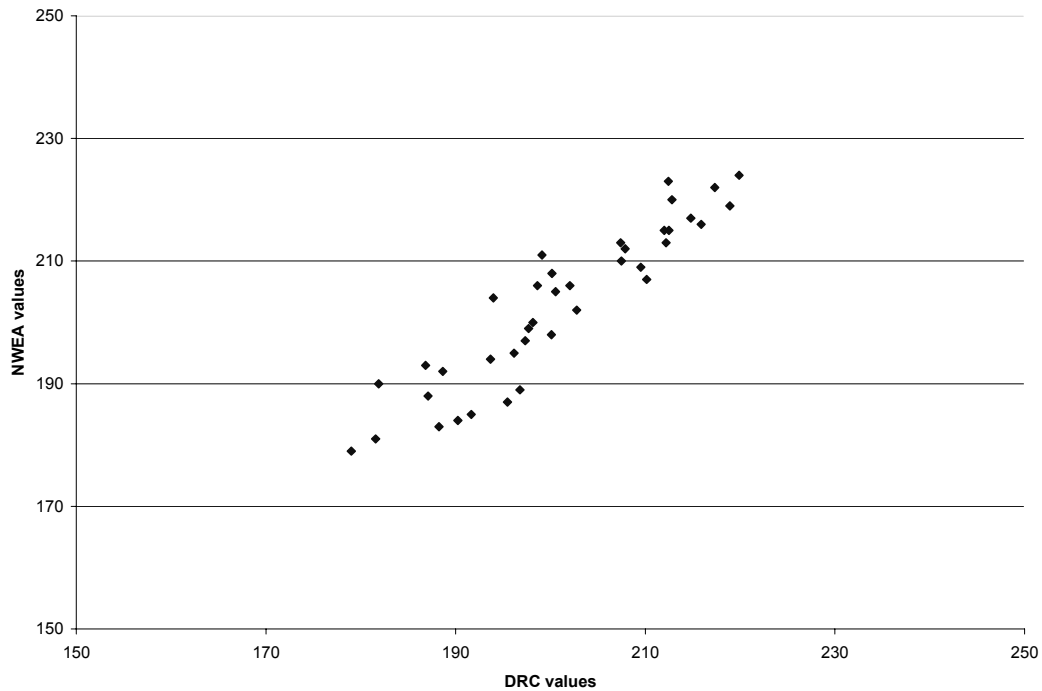
Independent Winsteps calibrations were computed for reading and mathematics at all grades; these values were compared to the scale values established by the former ISAT vendor. DRC's item calibration runs were performed without anchoring any values on the vertical scale. Therefore, the items were placed onto a common origin, determined by the mean of the NWEA scale score for items within grade.

Examples of the results of these calibrations are shown below for grade 4, one for reading, and one for mathematics. As a result of this study, Idaho-owned items with statistically significant differences were systematically eliminated from the July 2006 standard setting and Spring 2007 linking study. The examples below show plots of the two calibrations for the full set, as well as for two significance levels: alphas equal to .05 and .10 for a two-tailed test. Ultimately, Idaho-owned items were eliminated using  $\alpha=.05$ . The complete set of results may be found in Appendix A, for the full set only—without deletions.

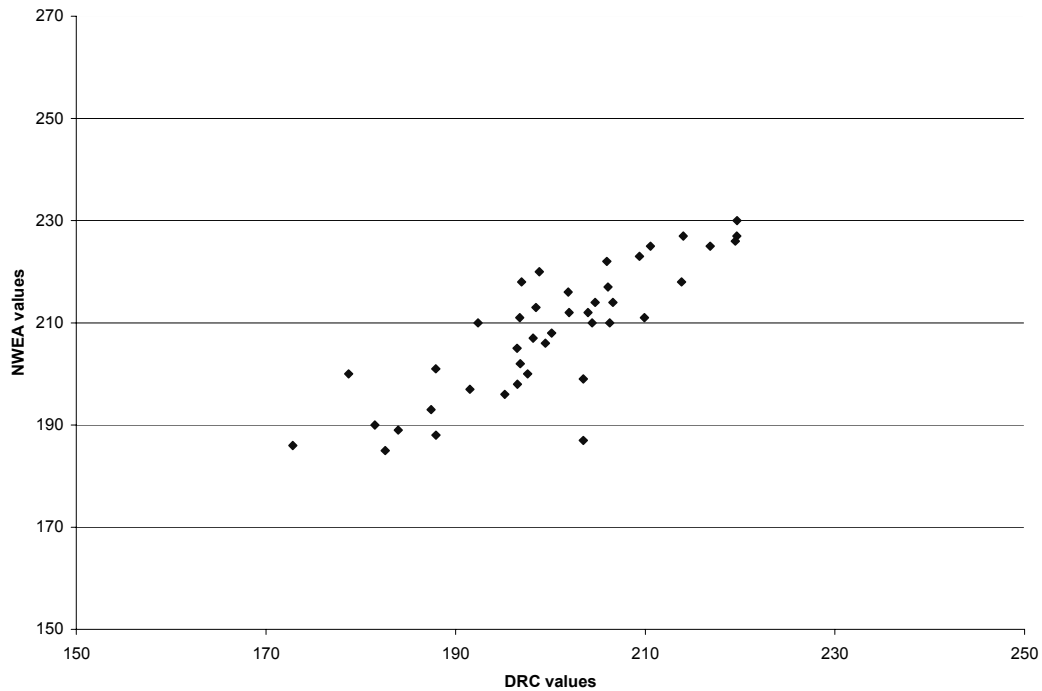
**Figure 4.5 Independent Item Calibrations—Scale Score Metric**  
**Grade 4 Reading—No Deletions**



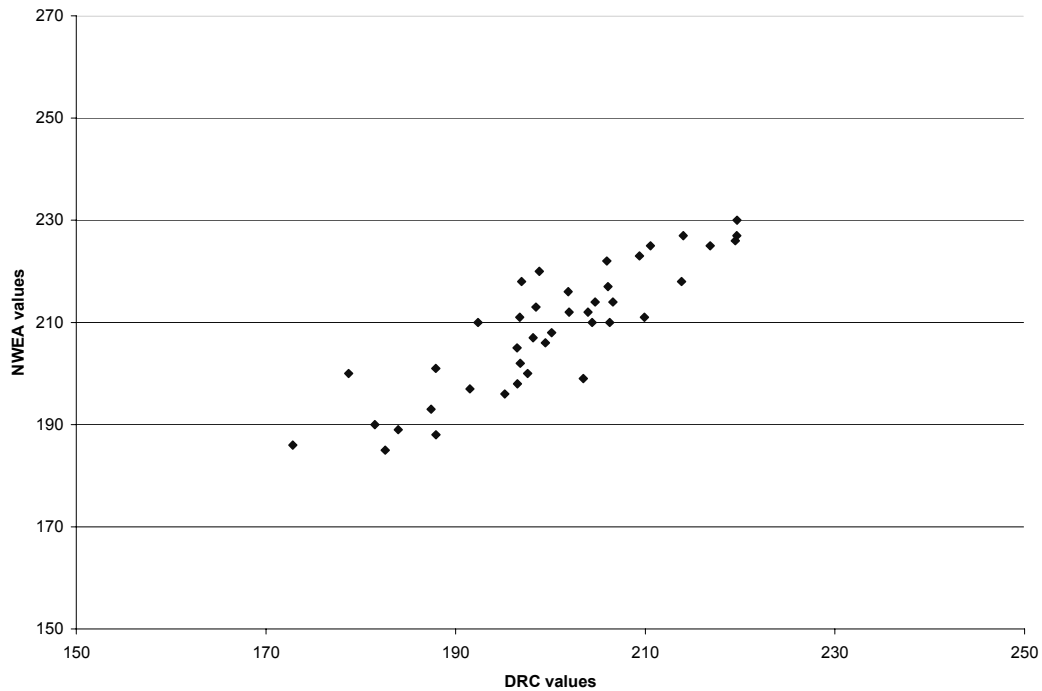
**Figure 4.6 Independent Item Calibrations—Scale Score Metric**  
**Grade 4 Reading—Deletions at  $z=1.960$**



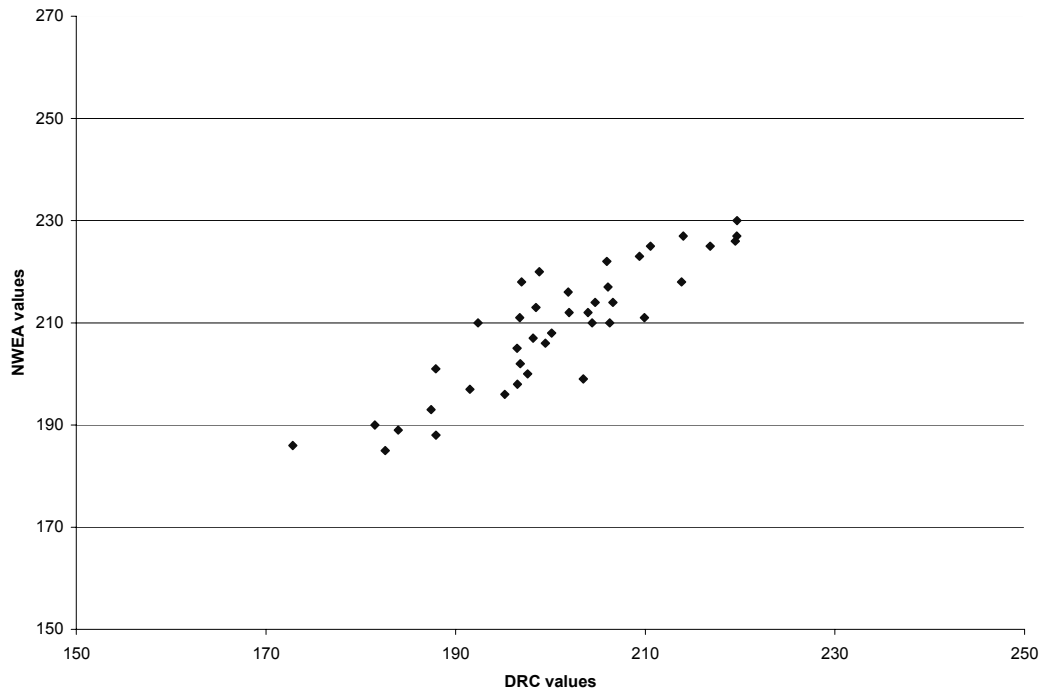
**Figure 4.7 Independent Item Calibrations—Scale Score Metric**  
**Grade 4 Mathematics—No Deletions**



**Figure 4.8 Independent Item Calibrations—Scale Score Metric**  
**Grade 4 Mathematics— Deletions at  $z=1.960$**



**Figure 4.9 Independent Item Calibrations—Scale Score Metric**  
**Grade 4 Mathematics— Deletions at  $z=1.645$**



## **Summary of Study One and Study Two Results**

These two studies showed that DRC was able to replicate and maintain the integrity of the assessment during transition. Upon presentation of the results to the OSBE and TAC, DRC proceeded with the selection of items for the July 2006 Standard Setting and the Spring 2007 Linking Study. The results from the latter study are described in detail in Chapter 14 *Linking*. For the former study, a separate technical report was generated and is presented here as Appendix W *Standard Setting and Standards Validation*.

### **4.2 Communication Plan**

DRC and CAL believe that there is great value that may be gathered from Idaho district and school personnel on various aspects of the ISAT administration. DRC and CAL communicated with districts and schools with the approval of the OSBE Contract Manager. Several focus group meetings were held in preparation for the Spring 2007 administration. Descriptions and summaries of five of the most significant of these focus groups are presented below.

#### **4.2.1 Technology Focus Group**

DRC, CAL, and OSBE met with personnel from several districts throughout the state in August 2006. The purpose of this meeting was to gather feedback on the functionality of the current testing system as well as solicit feedback on the CAL-developed testing engine. Several topics were discussed including data upload, student demographic data management, system requirements, tutorials and practice tests, and tools available in the testing engine. DRC and CAL used many of the suggestions from this meeting to assist in customizing the functionality of the CAL testing engine.

#### **4.2.2 Report Design Focus Group**

In October 2006, representatives from Blaine, Boise, and Meridian school districts met with DRC and OSBE to provide input on the design of the spring score reports. DRC provided report mockups for attendees to review. The individual student report, school reports, district reports, state reports, graduation summaries, and participation reports were discussed in detail at this meeting. The requirements for the state's raw data file were also discussed and a demonstration of DRC's Online Reporting System was presented. In addition, and as a result of the group discussion, class level reports and reporting of Alternate Assessment scores on the state's raw data file were requested by OSBE as an optional component. These reports were subsequently added via contract amendments.

#### **4.2.3 ISAT Overview Workshops**

In March 2007, DRC and OSBE conducted regional training sessions for district and school test coordinators at seven locations throughout the state. The purpose of these sessions was to introduce the DRC/CAL system and familiarize district and school personnel with the policies and procedures for the Spring 2007 ISAT administration. Topics covered at these sessions were system requirements, roles and responsibilities, test preparation, test security, test administration, accommodated materials, post test procedures, and important dates. The PowerPoint presentation presented at these sessions is provided in Appendix B. It is also available on the internet at <http://www.boardofed.idaho.gov/saa/documents/ISAT-RegTrainings2007.ppt>.

#### **4.2.4 Report Content, Use, and Interpretation Webex Sessions**

In May 2007, DRC held several Webex training sessions on the Spring 2007 ISAT score report content, use, and interpretation for district and school personnel throughout the state. Each district was encouraged to invite all personnel involved with the ISAT spring testing to participate in a Webex training session. Every district participated in at least one of the sessions. The purpose was to present the format and content of the DRC-generated score reports as well as to increase public awareness of the technical issues in measurement. The PowerPoint found in Appendix C was presented by DRC psychometric staff and participants had the opportunity to ask questions from a chat room setting following the session. Frequently asked questions based on these sessions were documented and posted to the ISAT website. The PowerPoint presentation was posted on the internet at [http://www.boardofed.idaho.gov/saa/documents/ISAT\\_ScoreRptPresentation\\_05-15-07.ppt](http://www.boardofed.idaho.gov/saa/documents/ISAT_ScoreRptPresentation_05-15-07.ppt). Frequently Asked Questions were posted on the internet at <http://www.boardofed.idaho.gov/saa/questions.asp>.

#### **4.2.5 Fall/Winter Computer Adaptive Testing Focus Group**

The primary purpose of the meeting was to gather feedback from a representative group of Idaho educators regarding the content and report format of the fall assessments, along with their use and interpretation. Eight participants were present for the entire day and two for half-days. Staff from DRC's Test Development and Psychometric Services presented to the group. Facilitation was provided by OSBE staff. Given the tight timelines for fall assessment implementation, the participants were told that their feedback would be used to assist DRC and the OSBE in designing the assessment for Fall 2008, and, if feasible, for Fall 2007.

DRC began the meeting with a presentation of the process and results of the standards validation meeting that was held the previous month. It was determined that the focus group discussion would be more productive if the participants understood more about standards validation as it had a direct impact on the spring results. The results of the standards validation meeting were used to make cut score recommendations that defined what percentage of Idaho students would be placed in the four achievement levels: Advanced, Proficient, Basic, and Below Basic.

A summary of the Fall/Winter CAT Focus Group is provided below which includes three presentations by DRC staff (Standards Validation, Content Considerations, and Psychometric Considerations) and the questions that were asked of the panel regarding the Fall/Winter Assessment. In addition, the reactions and suggestions of the panelists in making ongoing improvements to the ISAT program are provided in Appendix D.

## ***Standards Validation***

The standards validation report that was presented to the State Board for adoption was provided by DRC. Proficiency level cutpoints were determined for reading, mathematics, language usage, and science. The presentation included:

- The purpose and goals of the meeting
- A description of the method used to set cutpoints
- The number and background of the participants
- The process used to develop the proficiency level descriptors (PLDs)
- The involvement of Idaho educators in item development, item review, and the development of the PLDs
- The panelists' cutpoint recommendations based on their group ratings
- The cutpoint recommendations presented to the Board that were vertically articulated across grades (smoothed, but not made more or less strict overall) and how it was derived
- Results from Spring 2006 for comparison
- A summary of the participants' confidence in the process by means of a confidential survey filled out at the end of the meeting

## ***Content Considerations***

The presentation related to content considerations was provided by DRC. The presentation included:

- Purpose and goals of the focus group
- Importance of content alignment for the fall assessment
- Item cognitive complexity, areas of emphasis, and range of item difficulty
- Link between instruction and assessment
- Expectations for students in terms of skills and knowledge
- Multiple measures in decision making
- Summary of the differences between norm-references tests (NRTs) and criterion-based tests (CRTs or standards-based assessments) in terms of purpose, use, and interpretation
- Content considerations in building an appropriate item pool for computer adaptive testing (CAT)



### ***Psychometric Considerations***

The presentation related to psychometric considerations was provided by DRC. The presentation included:

- How a computer adaptive test algorithm works
- Psychometric goals of a CAT
- Report strategies
- Measures of growth

### ***Questions Asked of the Focus Group***

- What should the goal of the fall assessment be?
- What is an appropriate length of time for a fall assessment?
- What is an appropriate score to use, and interpretation of the score, for a fall assessment?
- How should the information from the reports be used?

## 5. New Item Development

The ISAT is designed to assess the content of the Idaho Content Standards in reading, mathematics, and language usage in grades 2–10 and in science in grades 5, 7, and 10. The Idaho Content Standards are divided into a three-tier hierarchical structure of 1) Content standards, 2) Goals, and 3) Objectives with content limits. The Idaho Content Standards are the overarching statements that describe in the very broadest terms what all students should know and be able to do as a result of the program. The Goals are the subset of the Idaho Content Standards setting forth a particular item of knowledge or skill(s) to be met by a student, and the Objectives (grade-level expectations) are the measurable demonstrations of a skill or item of knowledge required by a content goal or specifically what students should know and be able to do. The Idaho Objectives show progression from concrete to abstract over a sequence of years and serve as checkpoints that monitor progress toward students' mastery of the Idaho Content Standards. The Idaho Objectives for each content area and grade also serve to build understanding of the essential ideas across the years without a lot of repetition, and they attend to the prerequisite skills at a given grade level. Working in tandem with the Idaho Objectives are the contents limits which specify the skill or skills that are assessed and to what degree. In addition, they indicate whether an objective is assessed in the classroom only and not on the ISAT.

### ***Language Arts (Reading and Language Usage) Assessments Measures***

The language arts assessments of the ISAT are composed of items that address standards, goals, and objectives for grades 2–10 in two separate assessments, reading and language usage. The reading goals and objectives for each grade are distributed among two reporting categories: reading process and comprehension/interpretation.

The language usage goals and objectives for each grade are distributed among two reporting categories: writing process and writing components.

#### **Reading**

- **Reading Process:** vocabulary including context clues, affixes, synonyms, antonyms, use of headings and graphics
- **Comprehension/Interpretation:** main idea, relevant details, inference, conclusion of literary and expository texts, literary devices and figurative language, plot structure and text organization

#### **Language Usage**

- **Writing Process:** writing for a specific purpose and audience; selecting a main idea and supporting the main idea with relevant details; organizing ideas into paragraphs/essays which have a logical sequence and a clear beginning, middle, and an end; revising to clarify meaning by rearranging words or sentences, adding precise word choices, and/or eliminating irrelevant details; using a variety of sentences to create flow
- **Writing Components:** using correct sentence structure; editing for correct spelling, capitalization, punctuation, and grammar

## ***Mathematics Assessment Measures***

The mathematics assessment of the ISAT is composed of items that address standards, goals, and objectives for grades 2–10. The goals and objectives for each grade are distributed among five reporting categories—Number and Operations, Concepts and Principles of Measurement, Concepts and Language of Algebra and Functions, Principles of Geometry, and Data Analysis, Probability, and Statistics as outlined below:

- **Number and Operations:** understanding the properties of numbers; using numbers; computation; and estimation
- **Concepts and Principles of Measurement:** U.S. customary, metric, and time measurement, estimation, and conversion; two- and three-dimensional measurement computations; and use of rates, ratios, and proportions
- **Concepts and Language of Algebra and Functions:** using algebraic symbolism to represent mathematical relationships; using algebraic properties; evaluating algebraic expressions; solving equations; describing, extending, and using patterns
- **Principles of Geometry:** applying concepts of size, shape, and spatial relationships; applying concepts related to the Cartesian coordinate system; and graphing
- **Data Analysis, Probability, and Statistics:** analyzing, organizing, and displaying data in various formats; applying simple statistical measurements; determining and using probability; and making predictions based on data

## ***Science Assessment Measures***

The science assessment of the ISAT is composed of items that address standards, goals, and objectives for grades 5, 7, and 10. The goals and objectives for each grade are distributed among five reporting categories—Nature of Science; Physical Science; Biology; Earth and Space Systems; and Personal and Social Perspectives; Technology, as outlined below:

- **Nature of Science:** concepts of form and function, understanding of systems, processes of science, and interpretation of technical communication
- **Physical Science:** structure and properties of matter, molecules and atoms, and chemical reactions
- **Biology:** adaptations of plants and animals to their environment, cell forms and functions for all living things, and the relationship between matter and energy in living systems
- **Earth and Space Systems:** scientific theories of origin and changes in the Universe and Earth Systems, geo-chemical cycles and energy in the earth system
- **Personal and Social Perspectives; Technology:** relationships between science, society, natural resources, and technology.

## **5.1 Multiple-choice Items**

Multiple-choice items were used to assess the content objectives with content limits. For grade 2, the items require students to select a correct answer from three response choices with a single correct answer. For grades 3–10, students are required to select a correct answer from four response choices with a single correct answer. For grade 9 mathematics and the linking items for mathematics and science (all grades), students were required to select a correct answer from five response choices with a single correct answer. Each multiple-choice item is scored as right or wrong and has a value of one raw score point. Multiple-choice items are used to assess a variety of skill levels, from short-term recall of facts to problem solving. The selection of incorrect response choices, or distractors, by a student commonly can result from, but is not limited to, the following:

- misunderstood concepts
- incorrect logic, invalid application of an algorithm, or computation errors
- misinterpretation
- unsound reasoning
- superficial reading

## **5.2 Sampling of the Content Domain**

The reading, mathematics, language usage, and science content area reporting categories (or standards) are subdivided for specificity and eligible content or limits. Reading, mathematics, language usage, and science reporting categories are subdivided into one or more goals. Assessable and usable in statewide assessment, each respective goal is further subdivided into one or more objectives. Objectives that are not assessable and usable on the statewide assessment are identified for classroom assessment on the Item Specifications document per content and grade. Test items were not written for these classroom assessable objectives.

The number of objectives within each reporting category by content area and grade are listed in Tables 5.1–5.4.

**Table 5.1. Number of Assessable Reading Objectives  
By Reporting Category and Grade**

Standard	Grades							
	3	4	5	6	7	8	9	10
Reading Process	5	4	4	4	4	4	3	3
Comprehension/ Interpretation	9	13	13	12	10	11	9	8
<b>Total</b>	14	17	17	16	14	15	12	11

**Table 5.2. Number of Assessable Mathematics Objectives  
By Reporting Category and Grade**

Standard	Grades							
	3	4	5	6	7	8	9	10
Number and Operation	8	10	10	8	11	12	6	6
Concepts and Principals of Measurement	6	6	8	6	6	7	6	6
Concepts and Language of Algebra and Functions	8	7	8	9	9	11	7	7
Concepts and Principals of Geometry	3	4	6	4	5	6	8	8
Data Analysis, Probability, and Statistics	2	5	6	6	6	5	7	7
<b>Total</b>	27	22	38	33	37	41	34	34

**Table 5.3. Number of Assessable Language Usage Objectives  
By Reporting Category and Grade**

Standard	Grades							
	3	4	5	6	7	8	9	10
Writing Process	7	7	8	8	8	8	8	8
Writing Components	6	6	6	6	6	6	4	4
<b>Total</b>	13	13	14	14	14	14	12	12

**Table 5.4. Number of Assessable Science Objectives  
By Reporting Category and Grade**

Standard	Grades		
	5	7	10
Nature of Science	12	16	14
Physical Science	3	6	9
Biology	3	9	11
Earth and Space Systems	2	3	4
Personal and Social Perspectives; Technology	4	3	5
<b>Total</b>	24	37	43

## Test Development Process

Aligning the items to the content standards and content limits; determining the grade-level appropriateness (reading level/interest level, etc.); depth of knowledge; cognitive level; item/task level of complexity; estimated difficulty level; relevancy of context for each item; providing rationales for distractors; and determining style, accuracy, and correct terminology were major considerations in the item and test development process. *The Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) and the *Principles of Universal Design* (Thompson, Johnstone, & Thurlow, 2002) guided the following steps in the item and test development process.

1. Analyze the content standards and develop a preliminary test blueprint, including the standards to be assessed per content area and grade arranged by the relative importance of the content to be assessed (domains of content)
2. Develop preliminary item specifications and style guides
3. Develop preliminary proficiency-level descriptors
4. Select qualified item writers
5. Develop item-writing workshop training materials
6. Train test development specialists and item writers to write test questions
7. Write test questions that match the Idaho content standards and that are free of sensitivity concerns
8. Conduct and monitor internal test question reviews and quality processes
9. Prepare passages and test questions for review by the OSBE and committees of Idaho educators
10. Select and assemble test questions into test forms
11. Conduct field test
12. Review test questions and associated statistics after field testing
13. Select and assemble operational test forms (test construction)

### 5.3 Test Blueprints

The development of the content test blueprint began with the DRC test development specialists' in-depth analysis of the content standards and content limits, including discussion with representatives of the OSBE. Test development specialists then developed a preliminary test content test blueprint for reading, mathematics, language usage, and science at all grade levels. Each content test blueprint includes a list of all the standards, goals, and objectives to be assessed, organized by reporting categories. The preliminary test blueprint also contains the range of test items to be developed per goal and objective, with an indication of the relative importance of the content to be assessed. The preliminary content test blueprint for each content area and grade level was provided to OSBE and the Idaho TAC for review. DRC test development specialists revised each content test blueprint, based upon OSBE and the Idaho TAC feedback. The content test blueprint and content standards with limits were then used to guide the item and test development process.

## 5.4 Item Writing

The test questions are written by professional test question writers at DRC and professional writers from across the country who have successfully written test questions for large-scale assessment programs. All writers are experienced writers and teachers (current and former elementary and secondary) who have a great deal of specialized knowledge in the content area of their expertise. In addition, all writers possess good technical writing skills. The qualifications used to select writers include the following:

- A bachelor's degree or higher in reading, mathematics, language arts, science, curriculum and instruction, or related field
- In-depth understanding and knowledge of the special considerations involving the writing of criterion-referenced multiple-choice test questions, including an understanding of cognitive levels, estimated difficulty levels, grade-level appropriateness, readability, and sensitivity considerations

All item writers were provided with an in-depth training coupled with one-on-one writing sessions with DRC test development specialists and lead item writers. Prior to developing items for the ISAT the cadre of item writers were trained with regard to:

- Idaho content standards, goals, and objectives
- Cognitive levels, including Webb's depth of knowledge
- Principles of universal design
- Skill-specific and balanced test items for the grade level
- Contextual relevance
- Developmentally appropriate structure and content
- Item-writing technical quality issues
- Style considerations and item specifications approved by the OSBE

To ensure that all test items met the requirements of the approved content test blueprint and item specifications and were adequately distributed across subcategories and levels of difficulty, item writers were asked to document the following specific information as each item was written.

- **Alignment to the Idaho Standards, Goals, and Objectives:** There must be a high degree of match between a particular question and the objective it is intended to measure. Item writers were asked to clearly indicate what objective each item was measuring.
- **Estimated Difficulty Level:** Prior to field testing items, the item difficulties were not known, and writers could only make approximations as to how difficult an item might be. The estimated difficulty level was based upon the writer's own judgment as directly related to his or her classroom teaching and knowledge of the curriculum for a given content area and grade level. The purpose for indicating estimated difficulty levels as items were written was to help ensure that the pool of items prepared for review by Idaho educators and OSBE and subsequent field testing would include a range of difficulty (easy, medium, and challenging).

- **Appropriate Grade Level, Item Context, and Assumed Student Knowledge:** Item writers were asked to consider the conceptual and cognitive level of each item. They were asked to review each item to determine whether or not the item was measuring something that was important and could be successfully taught and learned in the classroom. In addition, item writers indicated the appropriate grade level of the item.
- **Multiple-choice Item Options:** Writers were instructed to make sure that each item had only one clearly correct answer. Item writers submitted the answer key with the item. All distractors were plausible choices that represented common errors and misconceptions in student reasoning.
- **Face Validity and Distribution of Complexity Levels:** Writers were instructed to write items to reflect various levels of cognitive complexity using Bloom et. al.'s *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain* (1956). As each item was written, the writer classified one of four cognition levels: recall, application, analysis, or evaluation for each item. The writers were instructed to write items so that the pool of items would represent a distribution of items across cognitive levels, as required by the test and item specifications.
- **Face Validity and Distribution of Items Based Upon Depth of Knowledge:** Writers were asked to classify the depth of knowledge of each item, using a model based on Norman Webb's work on depth of knowledge (Webb, 2002). Items were classified as one of four depths of knowledge categories: recall, skill/concept, strategic thinking, and extended thinking.
- **Readability:** For mathematics item development, writers were instructed to pay careful attention to the readability of each mathematics item to ensure that the focus was upon the concepts, not upon reading comprehension. As a result, the goal for each mathematics writer was to write items that were, to the greatest degree possible, independent of the assessment of reading. Content areas such as mathematics contain many content-specific vocabulary terms. These terms make it impossible to use the standard methods available for determining the reading level of test questions. Wherever it was practical and reasonable, every effort was made to keep the vocabulary one grade level below the tested grade level. Resources writers used to verify the vocabulary level were the *EDL Core Vocabularies* (Taylor et.al., 1989) and the *Children's Writer's Word Book* (Mogilner, 1992). In addition, every mathematics test question was taken before several different committees comprised of Idaho grade-level experts in the field of mathematics education. They reviewed each question from the perspective of the students they teach, and they determined the validity of the vocabulary used.
- **Curriculum-specific Issues:** All items were written to be curriculum independent with respect to both content and vocabulary. As items were written, writers were asked to document any specific curriculum issues.
- **Grammar and Structure for Item Stems and Item Options:** All items were written to meet technical quality, including correct grammar, syntax, and usage in all items, as well as parallel construction and structure of text associated with each multiple-choice item.



## **Editorial Review of Items**

After items were written, DRC test development specialists and editorial staff reviewed each item for item quality, making sure that the test items were in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for Idaho students. While there are many published guidelines for reviewing assessment items, the list below serves to summarize some of the more major considerations DRC test development specialists and editors followed when reviewing items to make sure they conformed to standard item quality for good, reliable, fair test questions.

### **Guidelines for Reviewing Assessment Items**

#### **A good item should:**

- have only one clear correct answer and contain answer choices that are reasonably parallel in length and structure.
- have a correctly assigned content code (item map).
- measure one main idea or problem.
- measure the objective or curriculum content standard it is designed to measure.
- be at the appropriate level of difficulty.
- be simple, direct, and free of ambiguity.
- make use of vocabulary and sentence structure that is appropriate to the grade level of the student being tested.
- be based on content that is accurate and current.
- when appropriate, contain stimulus material that are clear and concise and provide all of the information that is needed.
- when appropriate, contain graphics that are clearly labeled.
- contain answer choices that are plausible and reasonable in terms of the requirements of the question, as well as the students' level of knowledge.
- contain distractors that relate to the question in the same way and can be supported by a rationale.
- reflect current teaching and learning practices in the content area.
- be free of gender, ethnic, cultural, socioeconomic, and regional bias.

## **5.5 Passage Selection and Readability**

All reading items in the reading assessment were derived from a selection of literary and expository passages. Passages acquired were “authentic” in that they were culled from published materials or commissioned from experienced passage writers. To be used in the ISAT, approval to reprint published materials was secured from the publisher.

Passage finders and reading content specialists who have teaching experience at specific grade levels were given formal training on the specific requirements of the Idaho assessments. Passages were submitted to DRC’s reading test development team for screening and editing internally. The team screened and edited passages for:

- Interest and accuracy of information in a passage to a particular grade level
- Grade-level appropriateness of passage topic and vocabulary
- Rich passage content to support the development of high-quality test questions
- Bias, sensitivity, and fairness issues
- Readability considerations and concerns

Passages that survived this extensive screening process were prepared for review by the Idaho reading curriculum specialists for approval. Passages that were approved moved forward for the development of test questions.

The readability of a passage was a judgmental process made by Idaho educators, Idaho reading curriculum specialists, DRC’s reading content specialists, and other individuals who understand each particular grade level and children of a particular age group. In addition, formal readability programs were also used by DRC to provide a “snapshot” of a passage’s reading difficulty based on sentence structure, length of words, etc. All of this information, along with the classroom context and content appropriateness of a passage, was taken into consideration when placing a passage at a particular grade.

## **5.6 Universal Design**

As test questions were written and reviewed, the developers of the Idaho tests adhered closely to following guidelines for adhering to the principles of universal design.

Test questions measure what they are intended to measure. Item writing training included assuring that writers and reviewers have a clear understanding of Idaho’s content standards. During all phases of test development, test questions were presented with the content standard to ensure that each question measures what it is intended to measure. In certain types of test questions an additional skill may be necessary, such as a mathematics test requiring the student to read.

Test questions have concise and readable text. Linguistic demands of stimuli and test questions can interfere with a student's ability to demonstrate knowledge of the construct being assessed. Test questions were written to adhere to the following guidelines:

- Simple, clear, commonly used words are used whenever possible.
- Unnecessary words and extraneous text are omitted.
- Vocabulary and sentence complexity is appropriate for the grade level assessed.
- Technical terms and abbreviations are used only if they are related to the content being measured.
- Definitions and examples are clear and understandable.
- Idioms are avoided unless idiomatic speech is being assessed.

In addition, for the assessment to be fair to all students, test questions must also be clear in format. Images, pictures, and text that may not be necessary (e.g., sidebars, overlays, callout boxes, visual crowding, and shading) is distracting to students and is avoided. Pictures, illustrations, and graphics are used *only* if they provide essential information that contributes to the understanding of the constructs that test questions are intended to measure. The assessment must have an overall appearance that is clean and organized. In addition, minimal use of shading increases readability for students with visual difficulties. All test questions were presented in a way that allows for maximum readability for all students. For example, text presented in italics is far less legible and is read considerably more slowly than standard typeface, so it is used sparingly.

## **5.7 Depth of Knowledge**

Writers were asked to classify the depth of knowledge of each item, using a model based on Norman Webb's research on depth of knowledge. To facilitate discussion and execution of this task, writers used the Cognitive Level Comparison Matrix in Appendix E that presented side-by-side comparisons of the depth of knowledge categories of Bloom, Webb, and Porter. This at-a-glance summary provided definitions and examples at each category level. With Webb's work being the central focus, writers determined the language that best described Webb's four depth-of-knowledge categories: (1.0) recall; (2.0) basic application of skill/concept; (3.0) strategic thinking; and (4.0) extended thinking. By having access to the work of Bloom and of Porter, writers could formulate more accurate descriptions of the depth-of-knowledge levels for ISAT items.

## **5.8 Item Content Review**

All newly developed test items were submitted to content committees for review. The content committees consisted of Idaho educators from school districts throughout Idaho. The primary responsibility of the content committees was to evaluate items with regard to quality and content classification, including grade-level appropriateness, estimated difficulty, depth of knowledge, and source of challenge. They also suggested revisions and made recommendations for reclassification of items to different grade levels, if appropriate. The committees also reviewed the items for adherence to the principles of universal design, including language demand and issues of bias, fairness, and sensitivity. At the culmination of the item content review, all items that were presented to the committees for review were either accepted as presented or were revised.

The content review meeting was held January 15–17, 2007 in Boise, Idaho. All materials were submitted to the OSBE for approval prior to use with the review committees. The committee consisted of 62 educators, approximately 15 educators for each content area (mathematics, reading, language usage, and science) across the grade spans of 3 through 8 and 10. The committees represented the state of Idaho both geographically and demographically, including ELL and Special Education teachers. For demographics of the committee members, see Appendix Y. The meeting commenced with an overview of the test development process. Training was provided by DRC senior staff members using a PowerPoint presentation found in Appendix E. It included how to review items for both technical quality and content quality, including adherence to principles of universal design and depth-of-knowledge. In addition, reviewers were provided with a checklist of Universal Design Principles and a matrix comparing Bloom, Webb, and Porter cognitive levels. See Appendix E for these materials.

DRC test development specialists in reading, mathematics, language usage, and science facilitated the review of items. Committee members, grouped by grade span and content area, reviewed the items for quality and content, as well as for the following categories designated on the item tracking review form. An example of this form is found in Appendix F.

1. Standard/Goal/Objective Alignment (classified as Full, Partial, or No)
2. Complexity/Rigor (Bloom’s Taxonomy/Webb’s Depth of Knowledge/Porter’s Cognitive Demand)
3. Correct Answer (classified as Yes or No for MC items)
4. Graphics (classified as Yes or No, if graphics were acceptable)
5. Appropriate Language (classified as Yes or No)
6. Bias (classified as Yes or No)
7. Status/Overall Judgment (classified as Accept, Accept with Revisions, Revise or Rewrite, Drop)

Tables 5.5–5.8 show the number of items reviewed by content area/grade level and the number of items accepted by the item content review committees.

**Table 5.5. Number of Reading Items Reviewed and Accepted**

<b>January 2007 Item Review</b>		
<b>Grade</b>	<b>Number of Items Developed (reviewed at item review)</b>	<b>Number of Items Accepted</b>
3	160	158
4	180	174
5	183	182
6	157	156
7	154	154
8	161	161
10	176	176

**Table 5.6. Number of Mathematics Items Reviewed and Accepted**

January 2007 Item Review		
Grade	Number of Items Developed (reviewed at item review)	Number of Items Accepted
3	141	138
4	142	140
5	145	145
6	137	135
7	136	132
8	137	133
10	135	134

**Table 5.7. Number of Language Usage Items Reviewed and Accepted**

January 2007 Item Review		
Grade	Number of Items Developed (reviewed at item review)	Number of Items Accepted
3	150	149
4	150	150
5	150	149
6	150	150
7	150	147
8	150	147
10	150	146

**Table 5.8. Number of Science Items Reviewed and Accepted**

January 2007 Item Review		
Grade	Number of Items Developed (reviewed at item review)	Number of Items Accepted
5	184	171
7	177	171
10	145	141

## **5.9 Sensitivity and Bias Review**

Prior to field testing, all newly developed test items for grades 3–8 and 10 were also submitted to a Sensitivity Committee for review. This took place on January 15–16, 2007. The committee consisted of 14 educators. The committees represented the state of Idaho both geographically and demographically, including ELL and Special Education teachers. For demographics of the committee members see Appendix Y. The committee was trained by a DRC test development sensitivity specialist to review items for sensitivity issues using a PowerPoint presentation and Fairness in Testing Manual developed by DRC, found in Appendix G. The committee’s primary responsibility was to evaluate passages and items as to acceptability with regard to sensitivity issues. They also made recommendations for changes or deletion

of items in order to remove the area of concern. At the culmination of the sensitivity review, all items that were presented to the committee for review were either accepted as presented or were revised prior to the embedded field test.

All reading, mathematics, language usage, and science items were read by some of the committee members and some items were read by a cross section of members. Each member noted sensitivity comments on the Sensitivity Review Form (Appendix H). All comments were then compiled and the actions taken on these items were recorded by DRC.

### **5.10 Forms Approval Committee**

After DRC test development specialists constructed the operational forms, the forms were reviewed with a committee of Idaho educators on January 18–19, 2007, in Boise, Idaho. As indicated in Appendix Y, nineteen Idaho educators once again confirmed the items' alignment to standard, goal, and objective and approved the use of the items in the operational. Items that the committee members did not approve were replaced, and the form was then reviewed and approved by DRC psychometricians during the meeting. The committee members used a tracking sheet (Appendix I) to record the alignment of items to standard, goal, and objective.

### **5.11 Security**

Security was addressed by adhering to a strict set of procedures. Items in binders did not leave the meeting rooms and all were accounted for at the end of each day before attendees were dismissed. All attendees, with the exception of OSBE staff, were required to sign a Confidentiality Agreement (Appendix J). All materials not in use at any time were kept in secure meeting rooms. During lunch and breaks, if meeting rooms were unused, they were locked or closely monitored by DRC personnel. While not in use by DRC, the meeting rooms were locked and unavailable to anyone other than the Project Director or designee. Rooms were attended to only under strict supervision by DRC personnel. Secure materials that did not need to be retained after the meeting were deposited in secure containers, and their contents were shredded under supervision of a DRC employee.

## 5.12 Maintenance of the Item Bank

DRC maintains an item bank (IDEAS) that includes a record of all newly created items together with item data information from each field test of items. IDEAS also includes all data from the operational administration of the items, if the items appeared on an operational test form. For IDEAS maintenance, DRC:

- Updates the Idaho item bank after each administration
- Updates the Idaho item bank to include newly developed items
- Monitors the Idaho item bank to ensure an appropriate balance of items aligned with content standards, goals, and objectives
- Monitors item history statistics
- Monitors the Idaho item bank for an appropriate balance of DOK levels

### Summary of Field Test Items

The 2007 ISAT administration for reading, mathematics, language usage, and science was composed of one operational core form with ten embedded field test items for grades 3–8 and 10. The operational core form was randomized. The ISAT mathematics field test consisted of non-calculator and calculator items except for grades 3 and 4. Those students received non-calculator field test items only. Tables 5.9–5.12 provide the total number of field test items administered for reading, mathematics, language usage, and science.

**Table 5.9. Field Test Items for Reading**

<b>Grade</b>	<b>Total Number of Field Tested Items</b>
3	100
4	100
5	100
6	100
7	100
8	100
10	100

**Table 5.10. Field Test Items for Mathematics**

Grade	Multiple-Choice Items		Total Number of Field Tested Items
	Non-calculator	Calculator	
3	100	0	100
4	100	0	100
5	20	80	100
6	10	90	100
7	30	70	100
8	10	90	100
10	10	90	100

**Table 5.11. Field Test Items for Language Usage**

Grade	Total Number of Field Tested Items
3	100
4	100
5	100
6	100
7	100
8	100
10	100

**Table 5.12. Field Test Items for Science**

Grade	Total Number of Field Tested Items
5	100
7	100
10	100



## 6. Spring 2007 Embedded Field Test

### 6.1 Purpose

Field test items were administered to all grades and contents as part of the operational administration. The primary purpose of these items was to build the next year's operational forms. These items were presented within the core section so that student performance is as much like the operational test as possible. In this manner, the student was not aware of whether they are taking an operational item or a field test item. This ensures that psychometric analysis produces item difficulties that can be easily and directly translated onto the operational scale or metric. This process, commonly referred to as pre-equating, allowed for immediate scoring and reporting beginning in Spring 2008. As part of the embedded field test, approximately 100 items were administered in each grade and content.

### 6.2 Distractor Analysis

As with the operational assessment, the embedded field test consisted solely of multiple-choice items with four response options. Each multiple-choice item was scored as right or wrong. For each item, DRC calculated the percentage of students who selected each option and the option-total correlation for each of the options. Items with percent correct values between 37.5 and 95.0 are generally accepted as appropriate on statistical grounds, as are positive option-total correlations for the correct response and negative option-total correlations for the incorrect responses.

Table 6.1 shows an example for Grade 7 Language Arts. A complete set of distractor analyses for all grades and contents are presented in Appendix K.

**Table 6.1. Grade 7 Language Usage Field Test: Multiple Choice Distractor Analysis**

Item ID	Key	Response Options							
		A		B		C		D	
		P <sup>1</sup>	Pt-Bis	P <sup>1</sup>	Pt-Bis	P <sup>1</sup>	Pt-Bis	P <sup>1</sup>	Pt-Bis
511876	A	22.18	0.10	37.61	-0.05	28.62	0.03	11.59	-0.09
512094	B	14.31	-0.13	50.30	0.27	21.98	-0.14	13.40	-0.10
510873	A	62.01	0.43	6.09	-0.28	11.69	-0.17	20.21	-0.22
511483	A	44.32	0.31	11.20	-0.16	35.35	-0.14	9.13	-0.13
511485	D	3.78	-0.19	13.92	-0.18	22.53	-0.23	59.77	0.40
511486	A	56.34	0.20	10.73	-0.14	4.14	-0.20	28.80	-0.03
511346	B	17.74	-0.07	69.10	0.27	9.10	-0.20	4.07	-0.20
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
510876	B	12.34	-0.11	46.93	0.32	28.60	-0.15	12.13	-0.16
511431	C	5.61	-0.20	3.64	-0.21	86.10	0.39	4.65	-0.23
511488	D	13.62	-0.18	11.13	-0.21	22.32	-0.17	52.92	0.40
511339	A	54.07	0.43	25.51	-0.20	16.87	-0.25	3.55	-0.19
511430	B	7.40	-0.29	84.74	0.37	3.40	-0.18	4.46	-0.12
510877	D	21.81	-0.18	7.42	-0.28	7.42	-0.25	63.35	0.45
511619	C	15.20	-0.13	7.63	-0.30	72.05	0.39	5.12	-0.23
512814	A	45.57	0.29	18.09	-0.18	28.25	-0.10	8.09	-0.09
511433	A	66.15	0.37	16.92	-0.17	9.23	-0.31	7.69	-0.09

<sup>1</sup> Percentage of response options

Summary p-value information across all grades for each content is shown in Tables 6.2–6.5. The p-values are collapsed into blocks of deciles. Information in the table includes, for each grade, the p-value mean, median, standard deviation, minimum, and maximum. As one can see from the tables, most of the items fall into the 40 to 89 range, appropriate for a criterion-referenced assessment, and broad and deep enough (sufficient) to build strong core forms for the next assessment cycle. Science p-values are somewhat lower across the board, but well within an acceptable range; for mathematics, p-values are higher in grades 3 and 4 than in other grades, again, within acceptable ranges.

**Table 6.2. Reading Field Test: P-value Counts by Grade**

		Grade						
Range		3	4	5	6	7	8	10
	0-9	0	0	0	1	0	0	0
	10-19	1	0	1	2	2	0	2
	20-29	3	1	2	2	2	3	2
	30-39	6	7	5	1	4	4	4
	40-49	6	8	10	10	9	10	10
	50-59	15	9	17	13	12	19	16
	60-69	35	22	21	25	24	29	10
	70-79	25	30	19	27	27	26	29
	80-89	9	21	22	17	17	7	22
	90-99	0	2	3	2	3	2	5
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	62.79	67.63	65.62	65.94	66.00	63.03	67.08
	Median	65.17	71.15	67.40	68.85	69.11	65.01	71.22
	Std. Dev.	14.89	15.63	17.22	16.45	16.46	14.72	18.24
	Minimum	15.47	22.76	15.33	6.66	17.96	21.97	13.91
	Maximum	86.57	91.71	96.54	91.54	91.75	93.35	95.01

**Table 6.3. Mathematics Field Test: P-value Counts by Grade**

		Grade						
Range		3	4	5	6	7	8	10
	0-9	0	0	0	0	0	0	1
	10-19	0	2	0	0	1	1	2
	20-29	0	0	4	1	3	5	7
	30-39	1	2	5	4	4	7	10
	40-49	2	3	11	10	6	7	11
	50-59	4	10	9	17	17	22	19
	60-69	9	14	16	22	21	20	19
	70-79	23	25	23	18	24	20	13
	80-89	46	31	21	19	16	16	15
	90-99	15	13	11	9	8	2	3
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	80.18	74.33	68.37	67.73	67.53	62.66	58.38
	Median	84.07	76.63	72.27	68.98	69.06	64.86	59.39
	Std. Dev.	11.80	16.14	19.02	16.62	17.27	17.20	20.20
	Minimum	37.39	13.57	21.16	20.52	14.14	18.32	4.90
	Maximum	97.22	95.50	94.84	97.87	94.25	93.61	91.78

**Table 6.4. Language Usage Field Test: P-value Counts by Grade**

		Grade						
		3	4	5	6	7	8	10
<b>Range</b>	0-9	0	0	0	0	0	0	0
	10-19	0	1	0	1	1	2	0
	20-29	3	3	6	5	3	4	6
	30-39	5	7	8	9	4	7	10
	40-49	7	15	18	14	19	10	16
	50-59	25	14	19	20	21	22	11
	60-69	27	26	31	23	22	25	17
	70-79	21	21	10	16	16	16	27
	80-89	11	11	8	9	13	10	12
	90-99	1	2	0	3	1	4	1
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	63.38	61.97	57.99	58.95	60.52	60.79	61.07
	Median	65.18	64.64	59.72	60.18	61.79	63.99	64.97
	Std. Dev.	13.94	15.87	15.57	17.32	16.49	17.77	17.91
	Minimum	25.77	18.54	21.43	19.42	17.56	16.59	21.85
	Maximum	90.50	91.08	89.90	93.07	93.63	93.53	92.53

**Table 6.5. Science Field Test: P-value Counts by Grade**

		Grade		
		5	7	10
<b>Range</b>	0-9	1	0	0
	10-19	1	1	3
	20-29	9	12	10
	30-39	17	17	18
	40-49	20	20	19
	50-59	9	25	13
	60-69	14	13	21
	70-79	17	9	12
	80-89	9	3	4
	90-99	3	0	0
<b>Summary</b>	Total	100	100	100
	Mean	54.97	49.63	50.97
	Median	51.48	50.01	49.97
	Std. Dev.	20.55	16.17	17.56
	Minimum	9.75	15.88	18.32
	Maximum	94.16	87.66	87.42

Summary item-total information across all grades for each content is shown in Tables 6.6–6.9. The item-totals are collapsed into blocks of deciles. Information in the table includes, for each grade, the item-total correlation mean, median, standard deviation, minimum, and maximum. As one can see from the tables, most of the items fall into the .20 to .59 range, appropriate for a criterion-referenced assessment, with sufficient items in this range to build strong core forms for the next assessment cycle. Language usage and science have lower mean values than reading and mathematics across the grades, with science lower than language usage. This is an expected result given that science is a relatively new assessment in Idaho.

**Table 6.6. Reading Field Test: Item-Total Correlation Counts by Grade**

		<b>Grade</b>						
		<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>10</b>
<b>Range</b>	<0.0	1	0	1	1	2	0	0
	0.0 - 0.09	4	3	3	2	1	4	4
	0.10 - 0.19	4	2	3	4	3	5	9
	0.20 - 0.29	12	13	16	12	6	8	17
	0.30 - 0.39	20	25	29	35	30	41	35
	0.40 - 0.49	39	44	37	38	49	37	31
	0.50 - 0.59	19	13	11	8	9	5	4
	0.60 - 0.69	1	0	0	0	0	0	0
	0.70 - 0.79	0	0	0	0	0	0	0
	0.80 - 0.89	0	0	0	0	0	0	0
	0.90 - 0.99	0	0	0	0	0	0	0
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	0.39	0.39	0.38	0.37	0.39	0.36	0.34
	Median	0.42	0.42	0.39	0.38	0.41	0.38	0.36
	Std. Dev.	0.13	0.11	0.12	0.11	0.11	0.11	0.11
	Minimum	-0.01	0.06	-0.07	-0.23	-0.03	0.01	0.03
	Maximum	0.60	0.58	0.57	0.55	0.55	0.53	0.53

**Table 6.7. Mathematics Field Test: Item-Total Correlation Counts by Grade**

		<b>Grade</b>						
		<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>10</b>
<b>Range</b>	<0.0	0	1	0	0	1	0	2
	0.0 - 0.09	0	1	0	1	0	1	0
	0.10 - 0.19	2	5	4	1	6	10	6
	0.20 - 0.29	9	15	27	10	10	7	13
	0.30 - 0.39	34	38	32	33	38	29	30
	0.40 - 0.49	43	34	29	38	33	40	35
	0.50 - 0.59	12	6	8	17	11	13	13
	0.60 - 0.69	0	0	0	0	1	0	1
	0.70 - 0.79	0	0	0	0	0	0	0
	0.80 - 0.89	0	0	0	0	0	0	0
	0.90 - 0.99	0	0	0	0	0	0	0
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	0.40	0.36	0.36	0.41	0.38	0.38	0.38
	Median	0.41	0.37	0.36	0.41	0.39	0.41	0.40
	Std. Dev.	0.08	0.11	0.10	0.10	0.11	0.11	0.12
	Minimum	0.16	-0.20	0.12	0.02	-0.02	0.00	-0.11
	Maximum	0.56	0.59	0.59	0.59	0.60	0.57	0.65

**Table 6.8. Language Usage Field Test: Item-Total Correlation Counts by Grade**

		<b>Grade</b>						
		<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>10</b>
<b>Range</b>	<0.0	0	0	0	0	0	2	0
	0.0 - 0.09	1	4	2	3	1	1	1
	0.10 - 0.19	5	6	11	10	6	10	14
	0.20 - 0.29	11	14	17	20	20	14	24
	0.30 - 0.39	41	45	47	48	48	44	39
	0.40 - 0.49	40	31	23	17	24	28	21
	0.50 - 0.59	2	0	0	2	1	1	1
	0.60 - 0.69	0	0	0	0	0	0	0
	0.70 - 0.79	0	0	0	0	0	0	0
	0.80 - 0.89	0	0	0	0	0	0	0
	0.90 - 0.99	0	0	0	0	0	0	0
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	0.37	0.34	0.32	0.32	0.34	0.33	0.32
	Median	0.38	0.36	0.35	0.34	0.35	0.36	0.35
	Std. Dev.	0.09	0.10	0.10	0.10	0.09	0.11	0.10
	Minimum	0.06	0.02	0.05	0.02	0.08	-0.10	0.03
	Maximum	0.51	0.50	0.48	0.53	0.51	0.50	0.52

**Table 6.9. Science Field Test: Item-Total Correlation Counts by Grade**

		<b>Grade</b>		
		<b>5</b>	<b>7</b>	<b>10</b>
<b>Range</b>	<0.0	1	2	1
	0.0 - 0.09	10	6	11
	0.10 - 0.19	17	15	13
	0.20 - 0.29	28	29	21
	0.30 - 0.39	25	30	18
	0.40 - 0.49	19	16	27
	0.50 - 0.59	0	2	9
	0.60 - 0.69	0	0	0
	0.70 - 0.79	0	0	0
	0.80 - 0.89	0	0	0
	0.90 - 0.99	0	0	0
<b>Summary</b>	Total	100	100	100
	Mean	0.27	0.29	0.31
	Median	0.27	0.29	0.32
	Std. Dev.	0.12	0.13	0.15
	Minimum	-0.02	-0.07	-0.05
	Maximum	0.47	0.52	0.53

## 6.3 Item Analysis – Differential Item Functioning (DIF)

### Overview

Bias can present itself in a variety of ways in test items: through the language, the format, or the content. It can result from membership in a specific subpopulation or from factors correlated to the subpopulation. It can affect all members of the subpopulation, or it may affect only those in specific ranges of ability. Understanding how bias arises and how it presents itself has an impact on how best to detect and correct in test construction.

No statistical procedure should be used as a substitute for rigorous, hands-on reviews by content and bias specialists. The statistical results can help to frame and organize the review so the effort is concentrated on the most problematic cases; however, no items should be automatically rejected simply because a statistical method flagged them, nor automatically accepted because they were not flagged.

Statistical detection of item bias is at best an inexact science. There have been a variety of methods proposed for detecting bias, but no one statistic can be considered either necessary or sufficient. Different methods are more or less successful depending on the situation. No analysis can guarantee that a test is free of bias, but almost any thoughtful analysis will uncover the most flagrant problems.

A fundamental shortcoming of all of the statistical methods is that each is intrinsic to the test being evaluated. If a test, overall, is unbiased but with one or two biased items, any method will locate the problem items. If, however, all items on the test are consistently biased against a subpopulation, a statistical analysis of the items will not be able to separate bias effects from true differences in achievement.

### Mantel-Haenszel (M-H) Procedure for Differential Item Functioning

The *M-H* procedure (Mantel and Haenszel, 1959) for detecting differential item functioning is the most commonly used procedure in educational testing. It does not depend on the application or the fit of any specific measurement model. However, it does have some philosophical overlap with the Rasch model given that it uses total score to organize the analysis.

*Differential item functioning* is present when examinees of equal ability do not have the same probability of passing the item. If this inequity is associated with gender, ethnicity, or socioeconomic status, the item could be described as potentially biased.

The procedure as implemented by DRC contrasts a focal group with a reference group. While it makes no practical difference in the analysis which group is defined as focal and which as reference, the protected class or the group most apt to be disadvantaged by a biased measurement is typically defined as the focal group. The Mantel-Haenszel statistic (MH) for each item is computed from a two-by-two-by-eight contingency table. It has two groups (focal and reference), two outcomes (right or wrong) and eight ability groupings. The ability groups are defined by the octiles of the score distribution for the total examinee populations.

The basic MH statistic is a single degree of freedom chi-square that compares the observed number in each cell to the expected number. The expected counts are computed to ensure that the analysis is not confounded with differences in the achievement level of the two groups.

To assist in interpreting the results, the items are assigned severity classifications based on National Assessment of Educational Progress (NAEP) and Educational Testing Service (ETS) guidelines for dichotomous items (Allen, Carlson & Zelenak, 1999). Items classified as A+ or A- have no statistical indication of differential item functioning. Items classified as B+ or B- have some potential indication. Items classified as C+ or C- have strong potential evidence of a problem and should be carefully reviewed. As a convention, the plus sign indicates the item favors the focal group and a minus sign indicates the item favors the reference group. The formulas (and description of the variables) for the MH-statistic and the DIF categorization rules are included in Appendix L.

Below is an example of the item-level DIF statistics for Grade 3 Reading for the embedded field test portion of the Spring 2007 operational administration. The table includes an item identifier, the MH-delta, and the category designation for each pair-wise comparison. Items with insufficient numbers of students have blank cells. The complete set of item-level DIF statistics may be found in Appendix M.

**Table 6.10. Grade 3 Reading Field Test: Differential Item Functioning Analysis**

Item ID	Male/Female DIF NAEP		White/Hispanic DIF NAEP		No Aid/Lunch Aid DIF NAEP	
	MH delta	Cat.	MH delta	Cat.	MH delta	Cat.
511953	0.448	A+	-0.344	A-	-0.219	A-
511957	0.142	A+	0.234	A+	-0.036	A-
512282	0.438	A+	-0.432	A-	-0.389	A-
511960	0.747	A+	-0.244	A-	-0.164	A-
511955	0.030	A+	-0.455	A-	-0.853	A-
511959	1.020	B+	0.194	A+	-0.153	A-
511954	0.433	A+	-0.865	A-	-1.235	B-
511961	0.366	A+	-0.029	A-	-0.048	A-
511952	0.036	A+	0.851	A+	0.236	A+
511963	0.096	A+	-0.383	A-	-0.675	A-
511785	-0.632	A-			-0.184	A-
511782	-0.292	A-			-0.300	A-
511784	-0.293	A-			0.238	A+
511779	0.449	A+			-0.089	A-
511777	-0.605	A-			-0.092	A-
510888	0.678	A+			-0.054	A-
510887	-0.311	A-			-0.572	A-
510889	0.375	A+			-0.058	A-
509889	0.606	A+			-0.363	A-
509895	0.875	A+			0.299	A+
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
512324	-0.383	A-	-0.044	A-	0.508	A+
512331	-0.003	A-	-0.150	A-	0.127	A+
512329	-0.370	A-	-0.239	A-	-0.094	A-
512320	0.193	A+	0.296	A+	-0.573	A-
512326	0.370	A+	0.485	A+	-0.503	A-

Reference groups = Male, White, and No lunch aid.

Focal groups = Female, Hispanic or Latino, and Lunch Aid

Negative values reflect better reference group performance on the item and positive values reflect better focal group performance.

A summary of the counts of the number of items from each grade and content area that were assigned to each severity code is shown below in Tables 6.11–6.13. The tables include DIF summaries for male versus female, white versus Hispanic, and non lunch aid versus lunch aid.

**Table 6.11. DIF Counts by Content and Grade – Male/Female**

<b>Content</b>	<b>Grade</b>	<b>A+</b>	<b>A-</b>	<b>B+</b>	<b>B-</b>	<b>C+</b>	<b>C-</b>	<b>N/A</b>	<b>Total A</b>	<b>Total B</b>	<b>Total C</b>
Reading	3	63	35	2	0	0	0	0	98	2	0
Reading	4	59	33	5	1	2	0	0	92	6	2
Reading	5	45	32	18	2	3	0	0	77	20	3
Reading	6	60	29	5	5	0	1	0	89	10	1
Reading	7	49	36	9	3	3	0	0	85	12	3
Reading	8	40	53	2	4	0	1	0	93	6	1
Reading	10	57	31	8	4	0	0	0	88	12	0
Mathematics	3	47	46	3	2	1	1	0	93	5	2
Mathematics	4	51	37	5	5	2	0	0	88	10	2
Mathematics	5	59	29	5	6	0	1	0	88	11	1
Mathematics	6	53	39	2	5	1	0	0	92	7	1
Mathematics	7	57	31	9	0	3	0	0	88	9	3
Mathematics	8	48	38	8	3	3	0	0	86	11	3
Mathematics	10	49	40	4	5	2	0	0	89	9	2
Lang. Usage	3	47	52	0	0	0	0	1	99	0	0
Lang. Usage	4	63	34	1	1	0	0	1	97	2	0
Lang. Usage	5	68	29	0	1	0	0	2	97	1	0
Lang. Usage	6	67	29	3	0	0	0	1	96	3	0
Lang. Usage	7	60	35	2	1	0	1	1	95	3	1
Lang. Usage	8	65	27	3	3	1	0	1	92	6	1
Lang. Usage	10	62	36	1	0	0	0	1	98	1	0
Science	5	42	51	5	1	0	1	0	93	6	1
Science	7	58	38	0	3	0	1	0	96	3	1
Science	10	35	58	3	3	0	1	0	93	6	1



**Table 6.12. DIF Counts by Content and Grade – White/Hispanic or Latino**

<b>Content</b>	<b>Grade</b>	<b>A+</b>	<b>A-</b>	<b>B+</b>	<b>B-</b>	<b>C+</b>	<b>C-</b>	<b>N/A</b>	<b>Total A</b>	<b>Total B</b>	<b>Total C</b>
Reading	3	24	34	0	2	0	0	40	58	2	0
Reading	4	22	37	0	0	0	1	40	59	0	1
Reading	5	36	45	0	8	0	1	10	81	8	1
Reading	6	22	29	0	6	0	3	40	51	6	3
Reading	7	23	58	0	6	0	3	10	81	6	3
Reading	8	30	52	0	6	0	2	10	82	6	2
Reading	10	15	19	1	3	0	2	60	34	4	2
Mathematics	3	34	47	4	11	0	3	1	81	15	3
Mathematics	4	35	51	3	8	0	2	1	86	11	2
Mathematics	5	31	59	2	7	0	0	1	90	9	0
Mathematics	6	32	52	0	5	0	1	10	84	5	1
Mathematics	7	39	48	2	1	0	0	10	87	3	0
Mathematics	8	23	56	0	6	1	0	14	79	6	1
Mathematics	10	19	31	1	3	0	0	46	50	4	0
Lang. Usage	3	15	62	0	21	0	1	1	77	21	1
Lang. Usage	4	12	76	0	9	0	2	1	88	9	2
Lang. Usage	5	34	59	1	4	0	0	2	93	5	0
Lang. Usage	6	21	72	0	5	0	1	1	93	5	1
Lang. Usage	7	15	27	0	7	0	0	51	42	7	0
Lang. Usage	8	30	60	1	4	0	2	3	90	5	2
Lang. Usage	10	21	54	0	5	0	3	17	75	5	3
Science	5	22	51	0	12	0	2	13	73	12	2
Science	7	27	48	0	5	0	0	20	75	5	0
Science	10	7	6	0	1	0	0	86	13	1	0

**Table 6.13. DIF Counts by Content and Grade – No Lunch Aid/Lunch Aid**

<b>Content</b>	<b>Grade</b>	<b>A+</b>	<b>A-</b>	<b>B+</b>	<b>B-</b>	<b>C+</b>	<b>C-</b>	<b>N/A</b>	<b>Total A</b>	<b>Total B</b>	<b>Total C</b>
Reading	3	19	78	0	3	0	0	0	97	3	0
Reading	4	20	80	0	0	0	0	0	100	0	0
Reading	5	26	73	0	1	0	0	0	99	1	0
Reading	6	14	82	0	4	0	0	0	96	4	0
Reading	7	16	79	0	5	0	0	0	95	5	0
Reading	8	17	81	0	2	0	0	0	98	2	0
Reading	10	29	70	0	1	0	0	0	99	1	0
Mathematics	3	18	77	0	3	0	1	1	95	3	1
Mathematics	4	34	63	0	3	0	0	0	97	3	0
Mathematics	5	14	86	0	0	0	0	0	100	0	0
Mathematics	6	19	79	0	2	0	0	0	98	2	0
Mathematics	7	28	70	0	2	0	0	0	98	2	0
Mathematics	8	27	72	0	1	0	0	0	99	1	0
Mathematics	10	27	72	0	1	0	0	0	99	1	0
Lang. Usage	3	14	82	0	3	0	0	1	96	3	0
Lang. Usage	4	6	92	0	1	0	0	1	98	1	0
Lang. Usage	5	17	80	0	1	0	0	2	97	1	0
Lang. Usage	6	23	75	0	1	0	0	1	98	1	0
Lang. Usage	7	16	81	0	2	0	0	1	97	2	0
Lang. Usage	8	16	77	0	6	0	0	1	93	6	0
Lang. Usage	10	12	84	0	3	0	0	1	96	3	0
Science	5	22	75	0	3	0	0	0	97	3	0
Science	7	29	70	0	0	0	0	1	99	0	0
Science	10	38	60	0	1	0	0	1	98	1	0

## 6.4 Rasch Item Analysis

Scale scores for the ISAT were developed using the family of Rasch (1960) measurement models for scaling and equating. The advantage of using Rasch models in scaling is that all of the items measuring performance in a particular content area can be placed on a common difficulty scale, allowing the Rasch difficulty values for the individual items to be used in computing a Rasch logit for any raw score point on any test constructed from scaled items. This has the desirable properties of having the difficulty of the items and the student scores on a single scale. This allows for easier interpretation of the scores than is possible with more complex item response theory models that estimate more item parameters than difficulty (e.g., item discrimination and guessing).

Rather than percent correct, the Rasch model expresses item difficulty (and student proficiency) in units commonly referred to as logits. In the simplest case, a logit is a transformed  $p$ -value with the average  $p$ -value represented by a logit of zero. The logit metric has several mathematical advantages over  $p$ -values. It is an interval scale, meaning two items with logits of 0 and +1 are the same distance apart as items with logits of +3 and +4. Logits are independent of the ability distribution of the students taking a particular test. A specific form can have a mean logit of zero, whether the average  $p$ -value of the test is 0.8 or 0.3. The Rasch model also allows person measures and item measures to be placed on a common scale. This allows the comparison of person proficiency and item difficulty to determine the probability that a person will respond correctly to any given test item.

The standard Rasch calibration procedure sets the mean difficulty of the items on any unanchored calibration at zero. Any item with a  $p$ -value lower than the mean receives a positive logit and any item with a  $p$ -value higher than the mean receives a negative logit. Consequently, the logits for any calibration, whether it is a third grade reading test or a high school mathematics test, relate to an arbitrary origin defined by the average of item difficulties for that form. The average third grade reading item will have a logit of zero; the average high school mathematics item will have a logit of zero in unanchored calibrations. This logit scale applies to both item difficulties and student abilities.

The unconditional, joint maximum likelihood (UCON) estimation procedure estimates the person parameters (i.e., ability) simultaneously with the item parameters (i.e., difficulty). The UCON procedure was accomplished using WINSTEPS Version 3.63 (Linacre, 2006). This calibration software is commercially available and widely used in the testing industry and is considered the industry standard for Rasch calibration.

While the  $p$ -value summaries presented in section 6.2.1 provide useful within-grade information regarding the breadth and depth of the field test  $p$ -values, they do not provide useful information to compare across grades. Given that the ISAT is vertically scaled (a single interval-level scale that allows for comparisons across grades), a more useful way to review item summaries is to use the underlying vertical scale that preserves the same properties of the scale score metric that is used for final reporting. Tables 6.14–6.17 show the Rasch item difficulties in summary form across grades. These values are presented in logits, which for all intents and purposes are in the final scale metric save for a linear transformation. That is, multiplying the logits by 10 and adding 200 places the logit in the final scale score metric. Large negative logits represent easier items while large positive logits represent more difficult items.

The summary item difficulties are presented in groups of .5 logits from less than -3.0 to 6.0 logits and above. For reading, most of the items fall into the -2.0 to 1.49 range for grade 3 and slowly become more difficult as grades increase to the right, an expected result. The last column on the right for grade 10 shows that most of the items fall into the 0.0 to 3.49 range. At the bottom of each table is the mean,

median, standard deviation, minimum, and maximum for each grade. As can be expected, the values for all descriptive statistics except the standard deviation increase as the grades increase. The notable exception can be found in the reading table where the mean of the item difficulties for grade 10 is only slightly larger than the mean for grade 8. Without the aforementioned exception, the following tables for mathematics, language usage, and science follow the same pattern that is shown in reading. Individual Rasch item parameters and fit statistics may be found in Appendix N.

**Table 6.14. Reading Field Test: Item Difficulty Summary by Grade**

		Grade						
		3	4	5	6	7	8	10
<b>Range</b>	< -3.000	0	0	0	0	0	0	0
	-3.00 to -2.51	0	0	1	0	0	0	0
	-2.50 to -2.01	0	2	0	0	0	0	0
	-2.00 to -1.51	9	4	0	0	0	0	0
	-1.50 to -1.01	10	14	9	2	0	0	0
	-1.00 to -0.51	30	23	13	8	4	1	1
	-0.50 to -0.01	25	21	15	10	6	1	0
	0.00 to 0.49	14	13	16	20	14	2	11
	0.50 to 0.99	3	10	14	27	24	11	12
	1.00 to 1.49	6	10	16	17	22	17	14
	1.50 to 1.99	2	2	12	9	12	29	22
	2.00 to 2.49	1	1	1	2	9	18	9
	2.50 to 2.99	0	0	2	2	3	11	18
	3.00 to 3.49	0	0	1	2	4	7	6
	3.50 to 3.99	0	0	0	0	2	2	4
	4.00 to 4.49	0	0	0	0	0	1	1
	4.50 to 4.99	0	0	0	1	0	0	0
	5.00 to 5.49	0	0	0	0	0	0	2
	5.50 to 5.99	0	0	0	0	0	0	0
	>= 6.000	0	0	0	0	0	0	0
<b>Summary</b>	Total	100	100	100	100	100	100	100
	Mean	-0.411	-0.240	0.392	0.724	1.123	1.847	1.849
	Median	-0.479	-0.324	0.403	0.645	1.016	1.786	1.734
	Std. Dev.	0.852	0.911	1.066	0.987	0.990	0.834	1.124
	Minimum	-1.990	-2.095	-2.628	-1.209	-0.868	-0.534	-0.624
	Maximum	2.365	2.289	3.415	4.723	3.975	4.085	5.134

**Table 6.15. Mathematics Field Test: Item Difficulty Summary by Grade**

		Grade						
Range		3	4	5	6	7	8	10
	< -3.000	6	0	0	0	0	0	0
	-3.00 to -2.51	9	0	0	0	0	0	0
	-2.50 to -2.01	13	7	0	0	0	0	0
	-2.00 to -1.51	31	8	2	1	0	0	0
	-1.50 to -1.01	19	15	8	1	0	0	0
	-1.00 to -0.51	8	16	9	2	0	0	0
	-0.50 to -0.01	8	21	11	6	4	0	0
	0.00 to 0.49	5	13	17	10	6	1	0
	0.50 to 0.99	0	11	15	11	8	1	0
	1.00 to 1.49	1	5	12	16	21	6	0
	1.50 to 1.99	0	2	10	19	12	13	4
	2.00 to 2.49	0	0	10	18	18	12	9
	2.50 to 2.99	0	1	2	10	17	22	10
	3.00 to 3.49	0	1	4	5	6	23	12
	3.50 to 3.99	0	0	0	0	3	8	16
	4.00 to 4.49	0	0	0	1	4	8	21
	4.50 to 4.99	0	0	0	0	0	3	8
	5.00 to 5.49	0	0	0	0	1	2	9
	5.50 to 5.99	0	0	0	0	0	1	6
	>= 6.000	0	0	0	0	0	0	5
Summary	Total	100	100	100	100	100	100	100
	Mean	-1.568	-0.399	0.657	1.455	1.961	2.882	3.964
	Median	-1.750	-0.389	0.563	1.559	1.965	2.865	3.935
	Std. Dev.	0.928	1.093	1.227	1.103	1.105	1.024	1.236
	Minimum	-3.800	-2.428	-1.655	-1.869	-0.355	0.427	1.505
	Maximum	1.038	3.353	3.434	4.279	5.292	5.534	8.176

**Table 6.16. Language Usage Field Test: Item Difficulty Summary by Grade**

		Grade						
Range		3	4	5	6	7	8	10
	< -3.000	0	0	0	0	0	0	0
	-3.00 to -2.51	1	0	0	0	0	0	0
	-2.50 to -2.01	6	0	0	0	0	0	0
	-2.00 to -1.51	8	2	0	0	0	0	0
	-1.50 to -1.01	16	3	1	2	0	0	0
	-1.00 to -0.51	23	12	4	1	1	2	0
	-0.50 to -0.01	30	16	12	6	5	3	1
	0.00 to 0.49	7	27	11	11	8	4	0
	0.50 to 0.99	5	15	32	15	15	16	7
	1.00 to 1.49	4	15	14	24	19	14	16
	1.50 to 1.99	0	6	18	18	21	19	21
	2.00 to 2.49	0	3	5	10	21	20	17
	2.50 to 2.99	0	1	3	10	5	10	14
	3.00 to 3.49	0	0	0	3	3	8	12
	3.50 to 3.99	0	0	0	0	2	2	9
	4.00 to 4.49	0	0	0	0	0	2	3
	4.50 to 4.99	0	0	0	0	0	0	0
	5.00 to 5.49	0	0	0	0	0	0	0
	5.50 to 5.99	0	0	0	0	0	0	0
	>= 6.000	0	0	0	0	0	0	0
Summary	Total	100	100	100	100	100	100	100
	Mean	-0.655	0.362	0.897	1.318	1.507	1.756	2.243
	Median	-0.712	0.283	0.871	1.320	1.514	1.691	2.132
	Std. Dev.	0.816	0.866	0.833	0.970	0.917	1.002	0.951
	Minimum	-2.686	-1.670	-1.211	-1.197	-0.998	-0.731	-0.012
	Maximum	1.442	2.768	2.822	3.475	3.875	4.221	4.323

**Table 6.17. Science Field Test: Item Difficulty Summary by Grade**

		<b>Grade</b>		
		<b>5</b>	<b>7</b>	<b>10</b>
<b>Range</b>	< -3.000	0	0	0
	-3.00 to -2.51	0	0	0
	-2.50 to -2.01	3	0	0
	-2.00 to -1.51	2	0	0
	-1.50 to -1.01	7	1	0
	-1.00 to -0.51	17	0	0
	-0.50 to -0.01	8	9	2
	0.00 to 0.49	12	10	3
	0.50 to 0.99	18	20	10
	1.00 to 1.49	20	23	21
	1.50 to 1.99	8	17	11
	2.00 to 2.49	3	13	19
	2.50 to 2.99	1	6	17
	3.00 to 3.49	1	1	13
	3.50 to 3.99	0	0	4
	4.00 to 4.49	0	0	0
	4.50 to 4.99	0	0	0
	5.00 to 5.49	0	0	0
	5.50 to 5.99	0	0	0
	>= 6.000	0	0	0
<b>Summary</b>	Total	100	100	100
	Mean	0.338	1.201	1.990
	Median	0.565	1.194	2.045
	Std. Dev.	1.123	0.841	0.942
	Minimum	-2.370	-1.029	-0.205
	Maximum	3.154	3.151	3.859

## 7. Operational Test Forms Construction

The test forms construction process involved a very specific set of guidelines relative to the selection of multiple-choice items for the operational forms. DRC believes a key factor in forms construction is a solid understanding of the ISAT test/item specifications as well as blueprints and content limits for each content area as established by the OSBE, Idaho educators, and DRC test development specialists. By using a series of systematic steps to determine the technical quality of each item, including reviewing each item for alignment to standard, only items that reflect the full depth of knowledge and cognitive demands of the Idaho-adopted curriculum are used to construct the operational forms.

### **DRC's Steps in the Forms Construction Process**

1. DRC test development specialists review the test design blueprint, including the number of items per reporting category (standard) for each academic content area test.
2. DRC psychometricians analyze item data and flag potential problems.
3. DRC psychometricians provide test development specialists with an overview of the psychometric guidelines for forms construction.
4. DRC test development specialists and measurement experts receive training in forms construction, with a focus on requisite content validity and psychometric properties.
5. DRC test development specialists review all items in the operational pool and make an initial selection of items according to test blueprint guidelines, psychometric guidelines, and technical guidelines.
6. DRC test development specialists review initial linking item selection, following the guidelines for meeting psychometric and content technical quality from NWEA item pool.
7. Items selected for forms construction are reviewed by DRC senior-level test development specialists and measurement experts.
8. DRC test development specialists work with Idaho Forms Approval Committee to review and make replacements, if needed. If replacements are made, the form is reviewed again by DRC psychometricians.
9. Final sign-off takes place by DRC and OSBE, as required by the program.

### **7.1 2007 Operational Plan**

The 2007 ISAT in reading, mathematics, language usage, and science were comprised of one operational form randomized twenty times. All of the forms contained the core items that are identical for all students. In addition, each form included a set of ten field test items that were randomly embedded in fixed positions throughout the form. The embedded field test items primary purpose is to produce enough items to build the following year's operational (core) form. After completing the core and embedded field test items, all students took an extender set of fourteen items. These items served two purposes: linking items were used to link the Spring 2007 assessments to the operational scale and off-grade level items allowed for evaluation of the vertical scale.



- Grades 3–8 and 10 in reading and mathematics, and grades 5, 7, and 10 science forms were built using items field tested in December 2006 and approved by the forms approval meeting in January 2007. Grade 2 forms were built using newly developed pilot items. Therefore, students in grade 2 did not receive score reports. Grades 3–10 language usage and grade 9 reading, mathematics, and science operational forms were built using NWEA-developed items that were leased by the OSBE for use until February 2008.

Table 7.1 displays the design for the reading tests for grades 2–10 indicating the total number of items for each purpose. For example, students in grade 6 reading took a total of 66 items: 42 core items, 10 embedded field test items, and either 14 linking items or 14 off-grade level items.

**Table 7.1. 2007 Reading Operational Test Plan**

Grade	Multiple Choice Items		Total Operational Points	Off-grade Items*	Linking*	Total Number of Multiple Choice Items
	Core	FT				
2	35	0	35	0	0	35
3	40	10	40	14	14	64
4	40	10	40	14	14	64
5	42	10	42	14	14	66
6	42	10	42	14	14	66
7	44	10	44	14	14	68
8	44	10	44	14	14	68
9	45	0	45	14	14	59
10	45	10	45	14	14	69

\*Students were given either the off-grade or linking set of items.

The column entries for this table denote:

- the grade level
- the number of core multiple choice items
- the number of field test multiple choice items
- the total number of operational points
- the number of off-grade items
- the number of linking items
- the total number of multiple choice items

Table 7.2 displays the design for the mathematics tests for grades 2–10.

**Table 7.2. 2007 Operational Mathematics Test Plan**

Grade	Multiple Choice Items		Total Operational Points	Off-grade Items*	Linking*	Total Number of Multiple Choice Items
	Core	FT				
2	38	0	38	0	0	38
3	40	10	40	14	14	64
4	40	10	40	14	14	64
5	45	10	45	14	14	69
6	45	10	45	14	14	69
7	45	10	45	14	14	69
8	45	10	45	14	14	69
9	45	0	45	14	14	59
10	45	10	45	14	14	69

\*Students were given either the off-grade or linking set of items.

The column entries for this table denote:

- the grade level
- the number of core multiple choice items
- the number of field test multiple choice items
- the total number of operational points
- the number of off-grade items
- the number of linking items
- the total number of multiple choice items

Table 7.3 displays the design for the language usage tests for grades 2–10.

**Table 7.3. 2007 Operational Language Usage Test Plan**

Grade	Multiple Choice Items		Total Operational Points	Off-grade Items*	Linking*	Total Number of Multiple Choice Items
	Core	FT				
2	35	0	35	0	0	35
3	40	10	40	14	14	64
4	40	10	40	14	14	64
5	42	10	42	14	14	66
6	42	10	42	14	14	66
7	45	10	45	14	14	69
8	45	10	45	14	14	69
9	45	0	45	14	14	59
10	45	10	45	14	14	69

\*Students were given either the off-grade or linking set of items.

The column entries for this table denote:

- the grade level
- the number of core multiple choice items
- the number of field test multiple choice items
- the total number of operational points
- the number of off-grade items
- the number of linking items
- the total number of multiple choice items

Table 7.4 displays the design for the science tests for grades 5, 7, and 10.

**Table 7.4. 2007 Science Operational Test Plan**

Grade	Multiple Choice Items		Total Operational Points	Off-grade Items*	Linking*	Total Number of Multiple Choice Items
	Core	FT				
5	42	10	42	14	14	66
7	48	10	48	14	14	72
10	50	10	50	14	14	74

\*Students were given either the off-grade or linking set of items.

The column entries for this table denote:

- the grade level
- the number of core multiple choice items
- the number of field test multiple choice items
- the total number of operational points
- the number of off-grade items
- the number of linking items
- the total number of multiple choice items

## **7.2 Forms Construction Tools and Documentation**

DRC's content specialists collaborated with DRC's psychometricians and Idaho educators as part of the item selection and form construction process. The goal of this process was to select the core items for the operational forms and submit them to Idaho educators for their review and approval and then to OSBE for final review and approval.

Once the initial items were selected on each form, they were placed in DRC's proprietary electronic system for forms construction named FirstForm®.

### **FirstForm®**

DRC uses FirstForm® for both manual and automated forms construction. In manual mode, it provides maximal flexibility and ease of use for both content specialists and psychometricians. In automated mode, preliminary forms can be constructed based on a list of supplied constraints before they are reviewed by content specialists. The manual mode works best when the item pool is relatively thin, whereas the automated mode can significantly reduce the time required for building forms when the item pool is relatively large. For the ISAT, the forms were initially built manually, with all revisions made within FirstForm®. In either mode, formal electronic documentation is stored and may be downloaded and printed for formal sign-off or for any iteration in the construction process.

Throughout the process, content specialists from Test Development monitored the content distribution described by the test specifications and blueprints, ensuring that there are no potential problems related to developmental appropriateness, item cueing, or redundant content. DRC psychometricians examined the statistical quality of the preliminary form, paying specific attention to p-values, fit statistics, potential item bias, and key distribution. Any items found to be suspect from a statistical standpoint are reported to the content specialists for review and possible replacement. This process is repeated until a form satisfies the Idaho educators and the OSBE.

The items used to build the Spring 2007 operational ISAT test forms were selected from a pool of items garnered from a December 2006 field test administration. Therefore, the statistics associated with the items were not on the operational scale or metric. Based on this, no specific difficulty targets were used in building the Spring 2007 ISAT test forms. Instead, DRC test development specialists built the best possible core forms for content appropriateness and quality, matching the test blueprint. Once the forms were completed, DRC psychometric staff reviewed the forms for appropriate difficulty by grade and across grades for all four contents before giving final psychometric approval.

Figure 7.1 is an example of a forms construction target in FirstForm®.

**Figure 7.1**

**Target Definition**

☐ Use existing item set as target >>

☒ Enter a Rasch Logit to define a target curve >>

Rasch Logit Difficulty: .20

☐ Enter 3 IRT parameters to define a target curve >>

b: 0.0    a: 1.0    c: .20

b = location parameter  
a = slope parameter  
c = lower asymptote

☐ Enter a P-Value to be used as target difficulty >>

P-Value:

Cancel    Accept

Figure 7.2 is a screenshot that shows summary level data for both academic content and psychometrics. In the left table, the count and percents of each content code are displayed. On the right side of the tab, summaries are shown for item difficulties, p-values, item-total correlations, discrimination (as appropriate), and guessing (as appropriate). In addition, a distribution of the keys is shown.

Figure 7.2

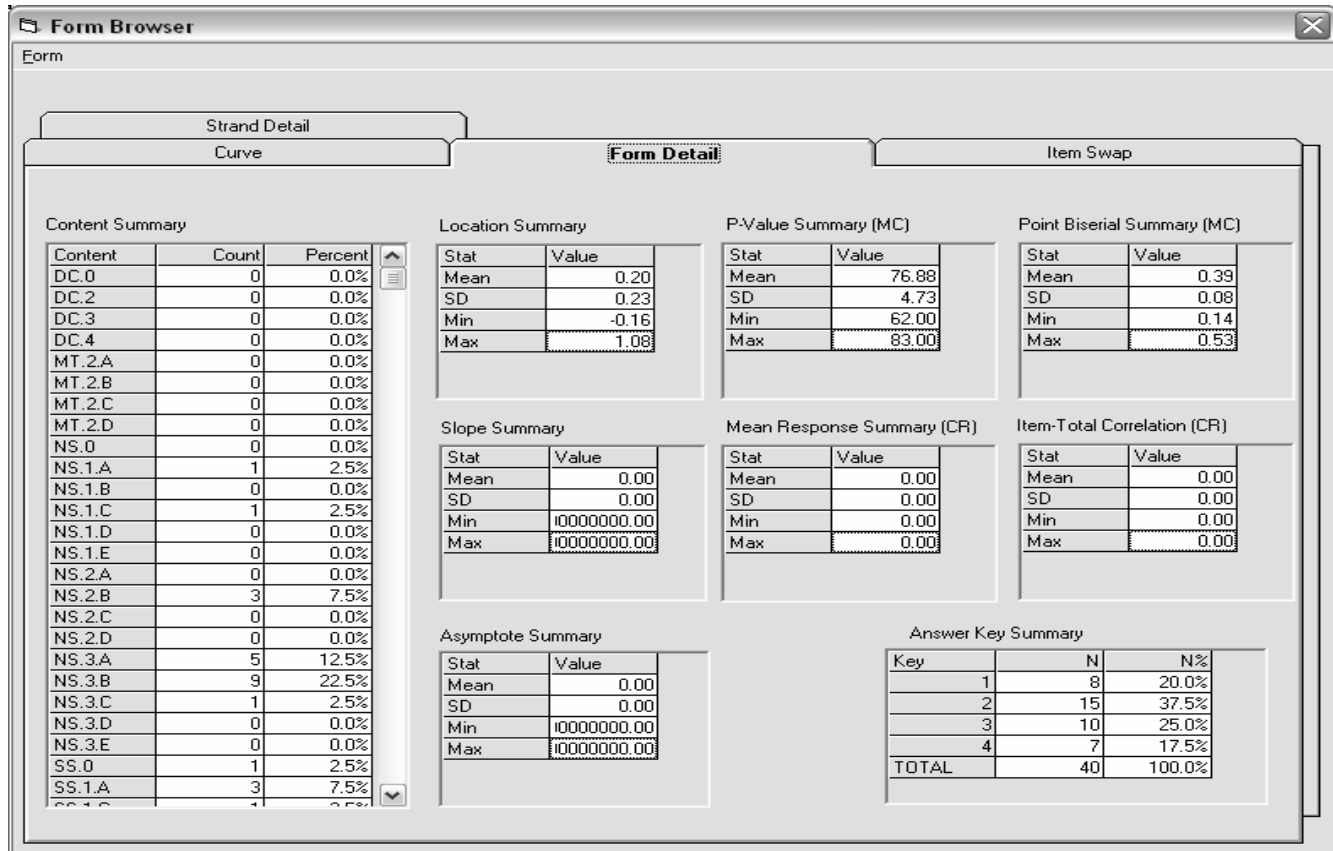


Figure 7.3 is a screenshot that shows the item swap feature. Notice that the current form is on the left and the residual (available) item pool is on the right. If the user chooses to replace a single item, they highlight the item to exclude on the left and hit the right arrow button, which moves the item from the current form to the residual pool. Then, a replacement is chosen from the right, highlighted, and moved to the current pool by clicking on the left arrow button. The “What If?” button can then be used to provide summary data for the new form or to print documentation.

**Figure 7.3**

**Form Browser**

Form

Strand Detail

Curve

Form Detail

**Item Swap**

Current Form

Item ID	Type	Content Code
03M99SS1A145...	MC	SS.1.A
03M99SS3D196...	MC	SS.3.D
03M99NS3B126...	MC	NS.3.B
03M99NS2B233...	MC	NS.2.B
03M99SS1A144...	MC	SS.1.A
03M99NS3B041...	MC	NS.3.B
03M99NS3B128...	MC	NS.3.B
03M99SS1D157...	MC	SS.1.D
03M99SS3A191...	MC	SS.3.A
03M99SS3C082...	MC	SS.3.C
03M99NS2B183...	MC	NS.2.B
03M99NS3A168...	MC	NS.3.A
03M99NS3B069...	MC	NS.3.B
03M99SS3C079...	MC	SS.3.C
03M99NS2B058...	MC	NS.2.B
03M99SS1D158...	MC	SS.1.D
03M00NS3A040...	MC	NS.3.A
03M00NS3A041...	MC	NS.3.A
03M00NS3A057...	MC	NS.3.A
03M00NS3A100...	MC	NS.3.A
03M00NS3B007...	MC	NS.3.B
03M00NS3B007...	MC	NS.3.B
03M00NS3B066...	MC	NS.3.B
03M00NS3B072...	MC	NS.3.B
03M00SS1A143...	MC	SS.1.A
03M00SS2B196...	MC	SS.2.B
03M00SS2C200...	MC	SS.2.C

Available Items

Item ID	Type	Content Code
03M99SS3D188...	MC	SS.3.D
03M99SS1C149...	MC	SS.1.C
03M99NS1A009...	MC	NS.1.A
03M99NS2B220...	MC	NS.2.B
03M99SS3C200...	MC	SS.3.C
03M99SS1A072...	MC	SS.1.A
03M99NS1A217...	MC	NS.1.A
03M99NS2C039...	MC	NS.2.C
03M99NS3B043...	MC	NS.3.B
03M99NS3C131...	MC	NS.3.C
03M99DC2095M...	MC	DC.2
03M99NS1A008...	MC	NS.1.A
03M99NS1A012...	MC	NS.1.A
03M99NS1A010...	MC	NS.1.A
03M99SS1D153...	MC	SS.1.D
03M99DC3099M...	MC	DC.3
03M99SS3C190...	MC	SS.3.C
03M99DC4101M...	MC	DC.4
03M99NS2A019...	MC	NS.2.A
03M99NS1A007...	MC	NS.1.A
03M99NS2A036...	MC	NS.2.A
03M99NS3B020...	MC	NS.3.B
03M99NS2A018...	MC	NS.2.A
03M99NS1A011...	MC	NS.1.A
03M99DC4100M...	MC	DC.4
03M99SS3C198...	MC	SS.3.C
03M99NS2B037...	MC	NS.2.B

>

>>

<

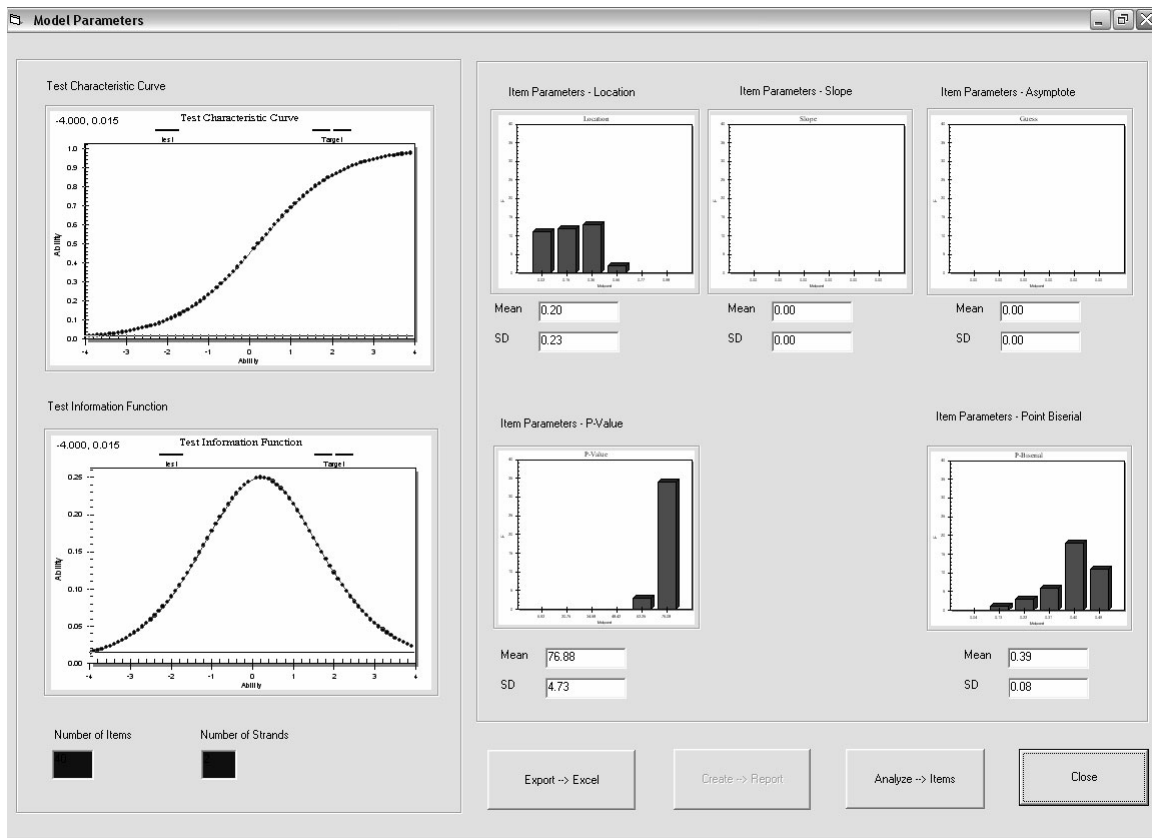
<<

What If?

Save

Figure 7.4 is a screenshot that shows the summary information for the current form. It displays the target test characteristic and information curves alongside the current forms, a histogram of the p-values and item-total correlations, and summary Rasch statistics. This screen, along with strand or subscore summaries, is included in the formal electronic documentation, along with a sign-off sheet for the lead content specialist, the lead psychometrician, and the client.

**Figure 7.4**





## 8. Test Administration

DRC, along with the online test delivery vendor, CAL, were the providers of the Spring 2007 ISAT. The tests were administered April 16–May 11, 2007 with make-up testing taking place May 14–18, 2007.

Reading, mathematics, and language usage tests were administered to students in grades 2–10. Science tests were administered to students in grades 5, 7, and 10. The items presented on the grade 2 tests were pilot items only and therefore, grade 2 students did not receive scores on their tests. Students were encouraged to review the ISAT Tutorial and take the ISAT Practice Tests prior to the administration of the online test. Test administrators were provided with detailed procedures on how to administer the tests. The Spring 2007 ISAT Test Administration Manual is provided in Appendix O. Over 600,000 tests were administered online for this administration.

During the Spring 2007 administration, a district feedback survey was distributed via email to all district test coordinators and school principals. A summary of the results of this survey is provided in Appendix P.

### 8.1 Sessions

Each content area was divided into several “parts” or test sessions.

**Table 8.1**

Grades	Content Area	# of Test Session Tickets	Part 1	Part 2	Part 3	Part 4
Grade 2	Reading, Language Usage and Mathematics	1	Pilot Study items			
Grades 3–10	Reading	2	NCLB* items	Extender** items		
Grades 3–10	Language Usage	2	NCLB* items	Extender** items		
Grades 5, 7, 10	Science	2	NCLB* items	Extender** items		
Grades 3, 4	Mathematics	3	NCLB* items	Extender** items	Extender** items	
Grades 5–10	Mathematics	4	Non-calculator NCLB* items	Calculator NCLB* items	Non-calculator Extender** items	Calculator Extender** items

Shaded areas are not part of that grade/content area.

\*No Child Left Behind (NCLB) federally mandated test

\*\*Extender section required by the Idaho State Board of Education

The students accessed each test session with a separate test session ticket. These tickets were printed by the test administrator or school test coordinator prior to testing. The test session ticket is a secure method to ensure students receive the correct grade level and content area test. During the spring administration, there were no students who received the incorrect test due to test session tickets.

## **8.2 Accommodations**

Paper-pencil, large print, and Braille versions of the test were available for those students whose Individualized Education Plan (IEP) required them. Districts ordered accommodated materials online and materials were delivered to districts 10 calendar days before the test window began. Students using the paper-pencil version marked their answers in a scannable answer document. Students using the large print version marked their answers directly in the test booklet and test administrators were responsible for transcribing the responses into a scannable answer document. Students using the Braille version used a Braille to mark their response and test administrators were responsible for transcribing the responses into a scannable answer document.

Districts were required to return all materials for arrival at DRC by May 22, 2007. Scores for students from districts who did not return materials by the deadline were not included in the score reports released on June 20, 2007. Approximately 86 accommodated versions of the test were administered.

## **8.3 Test Security**

### **Security of Test Content and Student Data**

Transmissions between the school districts' servers and the CAL application were encrypted using Secure Socket Layer (SSL) protocol. Only upon valid authentication (using Test Session Tickets) were the specific test and student data transmitted to students' computers. Once data were received and decrypted in local computers, they were loaded into the computer's memory, and no test or student data were ever written to hard disk. Once a student finished a test, data were completely deleted from memory. As soon as a student ended a test or if the test was interrupted (power failure, lost Internet signal, etc.), test data were completely removed from the local computers' memory.

The CAL application maintained total control of the student's computer during testing. Students taking the ISAT could not access other applications: printing, copying-and-pasting, screen capturing, keyboard shortcuts, and right-mouse clicks were all deactivated. During the Spring 2007 administration, there were no security breaches.

### **Security at the District and School Level**

District Test Coordinators (DTCs), School Test Coordinators, and Test Administrators shared the responsibility for ensuring that all test materials and student responses were handled securely and confidentially and in accordance with security procedures. The ISAT was to be administered by professional staff members who had been oriented in the proper test administration procedures for the ISAT used in the school.

A Test Security Agreement was provided in the Test Administration Manual and was required to be signed by all district and school personnel with access to the ISAT. DTCs were required to keep the agreements on file for two years. OSBE may audit these documents at any time.

All test booklets and answer documents, both used and unused, were inventoried by DRC upon receipt for scoring. There were a total of twenty-four missing secure materials from four districts for this administration.

## 9. Quality Control Processing

As students completed their tests online, CAL's system stored the responses. Because a Standards Validation was required after testing to validate the cut scores set in July 2006, immediate results were not available. Immediate results will be available for future administrations of the ISAT. CAL securely transmitted student response data to DRC daily for processing. DRC's student level processing system generated raw scores by applying the approved answer key to student responses. After cut scores were validated and approved by OSBE in May 2007, raw-to-scale conversion tables were applied to raw scores to determine scale scores and proficiency levels for generation of score reports.

The responses for students who used an accommodated version (paper-pencil, large print, or Braille) were marked on a scannable answer documents. Upon return to DRC, answer documents were scanned using OMR scanners. Throughout the scanning process, documents were checked for quality and scanning accuracy. After scanning was complete, the scan file containing student responses followed the same student level processing system that was used for the computer responses.

DRC's Quality Assurance department incorporated rigorous quality assurance activities throughout processing of student responses to ensure the highest level of quality and data integrity. Upon completion of the thorough data verification process, quality checks were performed on the data placement and report file formatting for each field on the reports. All reporting data elements were verified back to the live data file and the reporting processing rules. Additional quality cross-checks were performed to ensure accuracy and consistency across all reporting mediums for the assessment, including PDF files, CSV files, CDs, and the state's SQL database.

Similar programmatic routines were used to validate the data at the school, district, and state level. DRC's Senior Quality Assurance Analysts conducted a second review of each quality step to ensure the methodology, processes, and procedures were followed and to verify that the reports are approved for production generation.

To ensure the accuracy and reliability of the Report Delivery System, DRC's Software Quality Assurance Analysts validated that each Website page, link, and image displayed properly. They ensured that the system followed Graphical User Interface (GUI) standards and functioned as designed.

For this project, DRC adhered to the 33 Quality Control checkpoints for processing, scoring, and reporting described by the State Collaborative on Assessment and Student Standards (SCASS) on Technical Issues in Large Scale Assessments (TILSA). As added assurance, our Vice President of Quality conducted regular, formal, documented audits of our quality processes to ensure compliance to procedures. TILSA-approved quality checks were executed to verify that district and school content was correct and report data were 100-percent accurate.

## 10. Student Examinee Demographic Summary — Census

Student demographic breakdowns for the census population for the Spring 2007 ISAT are presented for each grade in Tables 10.1–10.8. Across contents, the numbers should be expected to be quite similar, as students are required to test all of these contents as part of NCLB (except for grade 9). Therefore, any differences in the counts and percentages across contents should be expected to be small. Counts and percents are presented for gender, ethnicity, lunch aid, limited English proficient, Title I, migrant status, gifted/talented, neglected/delinquent, homeless, home schooled, special education, and 504 plan. Enrollment counts, taken from the State Report of Participation Rates, may be found in Appendix X as a means to assess participation rates. That is, the number of students who were eligible for testing versus the number of students who actually tested.

**Table 10.1. Grade 3: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19872	100.0	19915	100.0	19878	100.0
<b>GENDER</b>						
Male	10251	51.6	10274	51.6	10253	51.6
Female	9621	48.4	9641	48.4	9625	48.4
<b>ETHNICITY</b>						
American Indian/Alaskan Native	292	1.5	291	1.5	292	1.5
Asian	233	1.2	234	1.2	234	1.2
Black/African American	246	1.2	251	1.3	247	1.2
Hawaiian/Other Pacific Islander	90	0.5	91	0.5	90	0.5
White	15716	79.1	15721	78.9	15726	79.1
Hispanic or Latino	2940	14.8	2973	14.9	2935	14.8
Other/Unknown	355	1.8	354	1.8	354	1.8
<b>FREE OR REDUCED LUNCH</b>						
No	10659	53.6	10669	53.6	10657	53.6
Yes	9213	46.4	9246	46.4	9221	46.4
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	18077	91.0	18075	90.8	18085	91.0
LEP	1395	7.0	1398	7.0	1394	7.0
LEP in first year of school	29	0.1	71	0.4	29	0.1
LEP exited in past 1 year	240	1.2	240	1.2	240	1.2
LEP exited in past 2 years	131	0.7	131	0.7	130	0.7
<b>TITLE I</b>						
No	12854	64.7	12865	64.6	12856	64.7
Yes	7018	35.3	7050	35.4	7022	35.3
<b>MIGRANT STATUS</b>						
No	19523	98.2	19560	98.2	19529	98.2
Yes	349	1.8	355	1.8	349	1.8

**Table 10.1 (continued). Grade 3: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>GIFTED AND TALENTED</b>						
No	19071	96.0	19115	96.0	19077	96.0
Yes	801	4.0	800	4.0	801	4.0
<b>NEGLECTED OR DELINQUENT</b>						
No	19872	100.0	19915	100.0	19878	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>HOMELESS</b>						
No	19805	99.7	19848	99.7	19812	99.7
Yes	67	0.3	67	0.3	66	0.3
<b>HOME SCHOOL</b>						
No	19872	100.0	19915	100.0	19878	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>						
No	17608	88.6	17646	88.6	17609	88.6
Yes	2030	10.2	2034	10.2	2035	10.2
SPE exited in past 1 year	170	0.9	170	0.9	170	0.9
SPE exited in past 2 years	64	0.3	65	0.3	64	0.3
<b>504 PLAN</b>						
No	19730	99.3	19772	99.3	19735	99.3
Yes	142	0.7	143	0.7	143	0.7

**Table 10.2. Grade 4: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19624	100.0	19680	100.0	19631	100.0
<b>GENDER</b>						
Male	10072	51.3	10102	51.3	10077	51.3
Female	9552	48.7	9578	48.7	9554	48.7
<b>ETHNICITY</b>						
American Indian/Alaskan Native	298	1.5	298	1.5	300	1.5
Asian	234	1.2	239	1.2	234	1.2
Black/African American	217	1.1	220	1.1	217	1.1
Hawaiian/Other Pacific Islander	72	0.4	72	0.4	72	0.4
White	15651	79.8	15662	79.6	15660	79.8
Hispanic or Latino	2782	14.2	2819	14.3	2780	14.2
Other/Unknown	370	1.9	370	1.9	368	1.9
<b>FREE OR REDUCED LUNCH</b>						
No	10740	54.7	10761	54.7	10746	54.7
Yes	8884	45.3	8919	45.3	8885	45.3
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	17836	90.9	17850	90.7	17847	90.9
LEP	1431	7.3	1432	7.3	1426	7.3
LEP in first year of school	16	0.1	57	0.3	17	0.1
LEP exited in past 1 year	225	1.1	225	1.1	225	1.1
LEP exited in past 2 years	116	0.6	116	0.6	116	0.6
<b>TITLE I</b>						
No	13236	67.4	13260	67.4	13242	67.5
Yes	6388	32.6	6420	32.6	6389	32.5
<b>MIGRANT STATUS</b>						
No	19317	98.4	19362	98.4	19327	98.5
Yes	307	1.6	318	1.6	304	1.5
<b>GIFTED AND TALENTED</b>						
No	18397	93.7	18453	93.8	18403	93.7
Yes	1227	6.3	1227	6.2	1228	6.3
<b>NEGLECTED OR DELINQUENT</b>						
No	19624	100.0	19680	100.0	19631	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>HOMELESS</b>						
No	19580	99.8	19636	99.8	19587	99.8
Yes	44	0.2	44	0.2	44	0.2
<b>HOME SCHOOL</b>						
No	19624	100.0	19680	100.0	19631	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>						
No	17398	88.7	17451	88.7	17404	88.7
Yes	1997	10.2	2000	10.2	1998	10.2
SPE exited in past 1 year	150	0.8	150	0.8	150	0.8
SPE exited in past 2 years	79	0.4	79	0.4	79	0.4
<b>504 PLAN</b>						
No	19448	99.1	19504	99.1	19455	99.1
Yes	176	0.9	176	0.9	176	0.9

**Table 10.3. Grade 5: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage		Science	
	N	%	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19880	100.0	19928	100.0	19887	100.0	19840	100.0
<b>GENDER</b>								
Male	10191	51.3	10213	51.2	10197	51.3	10164	51.2
Female	9689	48.7	9715	48.8	9690	48.7	9676	48.8
<b>ETHNICITY</b>								
American Indian/Alaskan Native	320	1.6	321	1.6	319	1.6	320	1.6
Asian	245	1.2	249	1.2	246	1.2	246	1.2
Black/African American	214	1.1	219	1.1	214	1.1	213	1.1
Hawaiian/Other Pacific Islander	81	0.4	81	0.4	81	0.4	81	0.4
White	15947	80.2	15953	80.1	15947	80.2	15908	80.2
Hispanic or Latino	2719	13.7	2749	13.8	2725	13.7	2721	13.7
Other/Unknown	354	1.8	356	1.8	355	1.8	351	1.8
<b>FREE OR REDUCED LUNCH</b>								
No	11137	56.0	11152	56.0	11142	56.0	11121	56.1
Yes	8743	44.0	8776	44.0	8745	44.0	8719	43.9
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18277	91.9	18280	91.7	18281	91.9	18236	91.9
LEP	1256	6.3	1257	6.3	1256	6.3	1254	6.3
LEP in first year of school	28	0.1	72	0.4	30	0.2	32	0.2
LEP exited in past 1 year	185	0.9	185	0.9	185	0.9	185	0.9
LEP exited in past 2 years	134	0.7	134	0.7	135	0.7	133	0.7
<b>TITLE I</b>								
No	14271	71.8	14294	71.7	14277	71.8	14255	71.8
Yes	5609	28.2	5634	28.3	5610	28.2	5585	28.2
<b>MIGRANT STATUS</b>								
No	19569	98.4	19615	98.4	19576	98.4	19530	98.4
Yes	311	1.6	313	1.6	311	1.6	310	1.6
<b>GIFTED AND TALENTED</b>								
No	18512	93.1	18560	93.1	18520	93.1	18474	93.1
Yes	1368	6.9	1368	6.9	1367	6.9	1366	6.9
<b>NEGLECTED OR DELINQUENT</b>								
No	19878	100.0	19926	100.0	19885	100.0	19838	100.0
Yes	2	0.0	2	0.0	2	0.0	2	0.0
<b>HOMELESS</b>								
No	19845	99.8	19893	99.8	19853	99.8	19806	99.8
Yes	35	0.2	35	0.2	34	0.2	34	0.2
<b>HOME SCHOOL</b>								
No	19880	100.0	19928	100.0	19887	100.0	19840	100.0
Yes	0	0.0	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>								
No	17639	88.7	17684	88.7	17643	88.7	17603	88.7
Yes	1987	10.0	1991	10.0	1991	10.0	1985	10.0
SPE exited in past 1 year	183	0.9	182	0.9	182	0.9	181	0.9
SPE exited in past 2 years	71	0.4	71	0.4	71	0.4	71	0.4
<b>504 PLAN</b>								
No	19660	98.9	19709	98.9	19668	98.9	19622	98.9
Yes	220	1.1	219	1.1	219	1.1	218	1.1

**Table 10.4. Grade 6: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19421	100.0	19477	100.0	19431	100.0
<b>GENDER</b>						
Male	9858	50.8	9890	50.8	9863	50.8
Female	9563	49.2	9587	49.2	9568	49.2
<b>ETHNICITY</b>						
American Indian/Alaskan Native	309	1.6	310	1.6	309	1.6
Asian	233	1.2	237	1.2	233	1.2
Black/African American	216	1.1	219	1.1	217	1.1
Hawaiian/Other Pacific Islander	81	0.4	81	0.4	81	0.4
White	15589	80.3	15607	80.1	15596	80.3
Hispanic or Latino	2662	13.7	2691	13.8	2662	13.7
Other/Unknown	331	1.7	332	1.7	333	1.7
<b>FREE OR REDUCED LUNCH</b>						
No	11126	57.3	11145	57.2	11130	57.3
Yes	8295	42.7	8332	42.8	8301	42.7
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	17929	92.3	17943	92.1	17939	92.3
LEP	1159	6.0	1161	6.0	1161	6.0
LEP in first year of school	14	0.1	55	0.3	13	0.1
LEP exited in past 1 year	203	1.0	202	1.0	203	1.0
LEP exited in past 2 years	116	0.6	116	0.6	115	0.6
<b>TITLE I</b>						
No	15441	79.5	15468	79.4	15445	79.5
Yes	3980	20.5	4009	20.6	3986	20.5
<b>MIGRANT STATUS</b>						
No	19129	98.5	19180	98.5	19139	98.5
Yes	292	1.5	297	1.5	292	1.5
<b>GIFTED AND TALENTED</b>						
No	18118	93.3	18174	93.3	18128	93.3
Yes	1303	6.7	1303	6.7	1303	6.7
<b>NEGLECTED OR DELINQUENT</b>						
No	19417	100.0	19473	100.0	19427	100.0
Yes	4	0.0	4	0.0	4	0.0
<b>HOMELESS</b>						
No	19392	99.9	19446	99.8	19402	99.9
Yes	29	0.1	31	0.2	29	0.1
<b>HOME SCHOOL</b>						
No	19421	100.0	19477	100.0	19431	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>						
No	17378	89.5	17427	89.5	17382	89.5
Yes	1803	9.3	1809	9.3	1808	9.3
SPE exited in past 1 year	166	0.9	167	0.9	167	0.9
SPE exited in past 2 years	74	0.4	74	0.4	74	0.4
<b>504 PLAN</b>						
No	19182	98.8	19237	98.8	19191	98.8
Yes	239	1.2	240	1.2	240	1.2



**Table 10.5. Grade 7: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage		Science	
	N	%	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19683	100.0	19714	100.0	19682	100.0	19587	100.0
<b>GENDER</b>								
Male	10182	51.7	10208	51.8	10186	51.8	10136	51.7
Female	9501	48.3	9506	48.2	9496	48.2	9451	48.3
<b>ETHNICITY</b>								
American Indian/Alaskan Native	322	1.6	322	1.6	321	1.6	312	1.6
Asian	251	1.3	255	1.3	253	1.3	252	1.3
Black/African American	210	1.1	207	1.1	209	1.1	207	1.1
Hawaiian/Other Pacific Islander	58	0.3	58	0.3	58	0.3	58	0.3
White	15901	80.8	15911	80.7	15901	80.8	15845	80.9
Hispanic or Latino	2612	13.3	2632	13.4	2612	13.3	2588	13.2
Other/Unknown	329	1.7	329	1.7	328	1.7	325	1.7
<b>FREE OR REDUCED LUNCH</b>								
No	11858	60.2	11858	60.2	11855	60.2	11805	60.3
Yes	7825	39.8	7856	39.8	7827	39.8	7782	39.7
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18245	92.7	18247	92.6	18241	92.7	18160	92.7
LEP	1140	5.8	1144	5.8	1143	5.8	1131	5.8
LEP in first year of school	32	0.2	58	0.3	33	0.2	32	0.2
LEP exited in past 1 year	139	0.7	138	0.7	139	0.7	138	0.7
LEP exited in past 2 years	127	0.6	127	0.6	126	0.6	126	0.6
<b>TITLE I</b>								
No	16764	85.2	16786	85.1	16769	85.2	16694	85.2
Yes	2919	14.8	2928	14.9	2913	14.8	2893	14.8
<b>MIGRANT STATUS</b>								
No	19414	98.6	19433	98.6	19412	98.6	19318	98.6
Yes	269	1.4	281	1.4	270	1.4	269	1.4
<b>GIFTED AND TALENTED</b>								
No	18602	94.5	18633	94.5	18602	94.5	18511	94.5
Yes	1081	5.5	1081	5.5	1080	5.5	1076	5.5
<b>NEGLECTED OR DELINQUENT</b>								
No	19672	99.9	19702	99.9	19670	99.9	19575	99.9
Yes	11	0.1	12	0.1	12	0.1	12	0.1
<b>HOMELESS</b>								
No	19650	99.8	19680	99.8	19648	99.8	19553	99.8
Yes	33	0.2	34	0.2	34	0.2	34	0.2
<b>HOME SCHOOL</b>								
No	19683	100.0	19714	100.0	19682	100.0	19587	100.0
Yes	0	0.0	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>								
No	17725	90.1	17756	90.1	17723	90.0	17658	90.2
Yes	1749	8.9	1748	8.9	1750	8.9	1723	8.8
SPE exited in past 1 year	143	0.7	144	0.7	143	0.7	140	0.7
SPE exited in past 2 years	66	0.3	66	0.3	66	0.3	66	0.3
<b>504 PLAN</b>								
No	19336	98.2	19367	98.2	19335	98.2	19243	98.2
Yes	347	1.8	347	1.8	347	1.8	344	1.8

**Table 10.6. Grade 8: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19840	100.0	19864	100.0	19821	100.0
<b>GENDER</b>						
Male	10246	51.6	10262	51.7	10238	51.7
Female	9594	48.4	9602	48.3	9583	48.3
<b>ETHNICITY</b>						
American Indian/Alaskan Native	310	1.6	310	1.6	309	1.6
Asian	238	1.2	239	1.2	238	1.2
Black/African American	179	0.9	179	0.9	179	0.9
Hawaiian/Other Pacific Islander	68	0.3	68	0.3	68	0.3
White	16093	81.1	16092	81.0	16077	81.1
Hispanic or Latino	2629	13.3	2651	13.3	2626	13.2
Other/Unknown	323	1.6	325	1.6	324	1.6
<b>FREE OR REDUCED LUNCH</b>						
No	12023	60.6	12032	60.6	12024	60.7
Yes	7817	39.4	7832	39.4	7797	39.3
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	18538	93.4	18525	93.3	18525	93.5
LEP	996	5.0	999	5.0	993	5.0
LEP in first year of school	33	0.2	69	0.3	32	0.2
LEP exited in past 1 year	145	0.7	144	0.7	144	0.7
LEP exited in past 2 years	128	0.6	127	0.6	127	0.6
<b>TITLE I</b>						
No	17405	87.7	17418	87.7	17391	87.7
Yes	2435	12.3	2446	12.3	2430	12.3
<b>MIGRANT STATUS</b>						
No	19582	98.7	19597	98.7	19564	98.7
Yes	258	1.3	267	1.3	257	1.3
<b>GIFTED AND TALENTED</b>						
No	18716	94.3	18740	94.3	18697	94.3
Yes	1124	5.7	1124	5.7	1124	5.7
<b>NEGLECTED OR DELINQUENT</b>						
No	19815	99.9	19838	99.9	19795	99.9
Yes	25	0.1	26	0.1	26	0.1
<b>HOMELESS</b>						
No	19812	99.9	19835	99.9	19791	99.8
Yes	28	0.1	29	0.1	30	0.2
<b>HOME SCHOOL</b>						
No	19840	100.0	19864	100.0	19821	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>						
No	17952	90.5	17980	90.5	17937	90.5
Yes	1717	8.7	1712	8.6	1712	8.6
SPE exited in past 1 year	105	0.5	107	0.5	106	0.5
SPE exited in past 2 years	66	0.3	65	0.3	66	0.3
<b>504 PLAN</b>						
No	19484	98.2	19508	98.2	19464	98.2
Yes	356	1.8	356	1.8	357	1.8

**Table 10.7. Grade 9: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage	
	N	%	N	%	N	%
<b>ALL STUDENTS</b>	20544	100.0	20611	100.0	20523	100.0
<b>GENDER</b>						
Male	10645	51.6	10676	51.8	10637	51.8
Female	9899	48.4	9935	48.2	9886	48.2
<b>ETHNICITY</b>						
American Indian/Alaskan Native	322	1.5	321	1.6	321	1.6
Asian	243	1.2	252	1.2	242	1.2
Black/African American	193	1.2	201	1.0	194	0.9
Hawaiian/Other Pacific Islander	65	0.5	65	0.3	64	0.3
White	16773	79.1	16789	81.5	16763	81.7
Hispanic or Latino	2585	14.8	2619	12.7	2579	12.6
Other/Unknown	363	1.8	364	1.8	360	1.8
<b>FREE OR REDUCED LUNCH</b>						
No	12877	53.6	12904	62.6	12870	62.7
Yes	7667	46.4	7707	37.4	7653	37.3
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	19361	91.0	19366	94.0	19343	94.3
LEP	952	7.0	958	4.6	950	4.6
LEP in first year of school	31	0.1	87	0.4	30	0.1
LEP exited in past 1 year	113	1.2	113	0.5	113	0.6
LEP exited in past 2 years	87	0.7	87	0.4	87	0.4
<b>TITLE I</b>						
No	19119	64.7	19185	93.1	19108	93.1
Yes	1425	35.3	1426	6.9	1415	6.9
<b>MIGRANT STATUS</b>						
No	20281	98.2	20341	98.7	20261	98.7
Yes	263	1.8	270	1.3	262	1.3
<b>GIFTED AND TALENTED</b>						
No	19603	96.0	19668	95.4	19581	95.4
Yes	941	4.0	943	4.6	942	4.6
<b>NEGLECTED OR DELINQUENT</b>						
No	20477	100.0	20542	99.7	20454	99.7
Yes	67	0.0	69	0.3	69	0.3
<b>HOMELESS</b>						
No	20522	99.7	20589	99.9	20501	99.9
Yes	22	0.3	22	0.1	22	0.1
<b>HOME SCHOOL</b>						
No	20544	100.0	20611	100.0	20523	100.0
Yes	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>						
No	18669	88.6	18739	90.9	18649	90.9
Yes	1735	10.2	1732	8.4	1735	8.5
SPE exited in past 1 year	92	0.9	92	0.4	92	0.4
SPE exited in past 2 years	48	0.3	48	0.2	47	0.2
<b>504 PLAN</b>						
No	20176	99.3	20244	98.2	20156	98.2
Yes	368	0.7	367	1.8	367	1.8

**Table 10.8. Grade 10: Summary of Student Demographics**

	Reading		Mathematics		Lang. Usage		Science	
	N	%	N	%	N	%	N	%
<b>ALL STUDENTS</b>	19092	100.0	19131	100.0	19101	100.0	18769	100.0
<b>GENDER</b>								
Male	9799	51.3	9816	51.3	9796	51.3	9618	51.2
Female	9293	48.7	9315	48.7	9305	48.7	9151	48.8
<b>ETHNICITY</b>								
American Indian/Alaskan Native	249	1.3	246	1.3	249	1.3	225	1.2
Asian	219	1.1	227	1.2	219	1.1	216	1.2
Black/African American	178	0.9	178	0.9	175	0.9	168	0.9
Hawaiian/Other Pacific Islander	80	0.4	80	0.4	80	0.4	77	0.4
White	15883	83.2	15881	83.0	15891	83.2	15640	83.3
Hispanic or Latino	2141	11.2	2175	11.4	2144	11.2	2112	11.3
Other/Unknown	342	1.8	344	1.8	343	1.8	331	1.8
<b>FREE OR REDUCED LUNCH</b>								
No	12959	67.9	12970	67.8	12965	67.9	12737	67.9
Yes	6133	32.1	6161	32.2	6136	32.1	6032	32.1
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18095	94.8	18100	94.6	18104	94.8	17791	94.8
LEP	810	4.2	808	4.2	811	4.2	792	4.2
LEP in first year of school	17	0.1	52	0.3	15	0.1	16	0.1
LEP exited in past 1 year	103	0.5	103	0.5	103	0.5	102	0.5
LEP exited in past 2 years	67	0.4	68	0.4	68	0.4	68	0.4
<b>TITLE I</b>								
No	18380	96.3	18414	96.3	18389	96.3	18064	96.2
Yes	712	3.7	717	3.7	712	3.7	705	3.8
<b>MIGRANT STATUS</b>								
No	18881	98.9	18914	98.9	18889	98.9	18559	98.9
Yes	211	1.1	217	1.1	212	1.1	210	1.1
<b>GIFTED AND TALENTED</b>								
No	18312	95.9	18352	95.9	18322	95.9	17997	95.9
Yes	780	4.1	779	4.1	779	4.1	772	4.1
<b>NEGLECTED OR DELINQUENT</b>								
No	19008	99.6	19047	99.6	19018	99.6	18689	99.6
Yes	84	0.4	84	0.4	83	0.4	80	0.4
<b>HOMELESS</b>								
No	19074	99.9	19113	99.9	19083	99.9	18752	99.9
Yes	18	0.1	18	0.1	18	0.1	17	0.1
<b>HOME SCHOOL</b>								
No	19092	100.0	19131	100.0	19101	100.0	18769	100.0
Yes	0	0.0	0	0.0	0	0.0	0	0.0
<b>SPECIAL EDUCATION</b>								
No	17443	91.4	17479	91.4	17449	91.4	17198	91.6
Yes	1550	8.1	1552	8.1	1551	8.1	1472	7.8
SPE exited in past 1 year	67	0.4	67	0.4	68	0.4	66	0.4
SPE exited in past 2 years	32	0.2	33	0.2	33	0.2	33	0.2
<b>504 PLAN</b>								
No	18765	98.3	18803	98.3	18775	98.3	18452	98.3
Yes	327	1.7	328	1.7	326	1.7	317	1.7

## 11. Operational Item and Form Summary

### 11.1 Distribution of P-values and Item-Total Correlations by Grade

Summary p-value information across all grades for each content is shown in Tables 11.1–11.4 for each operational core form. The p-values are collapsed into blocks of deciles. Information in the table includes, for each grade, the p-value mean, median, standard deviation, minimum, and maximum. Most of the items fall into the 40 to 89 range, appropriate for a criterion-referenced assessment. The mean p-value was somewhat lower for reading in grade 4 than for the other grades, somewhat higher in mathematics for grade 3, higher in language arts for grade 3 and lower for grades 9 and 10, and in science, lower across the board than the other contents.

Summary item-total correlation information across all grades for each content is shown in Tables 11.5–11.8 for each operational core form. The item-totals are collapsed into blocks of deciles. Information in the table includes, for each grade, the item-total mean, median, standard deviation, minimum, and maximum. As can be seen from the tables, most of the items fall into the .20 to .59 range, appropriate for a criterion-referenced assessment. Language usage and science have lower mean values than reading and mathematics across the grades, with science lower than language usage. This may be an expected result given that science is a relatively new assessment in Idaho. Note that, in language usage, the core test consisted of items developed by the former ISAT vendor.

The multiple-choice distractor analysis for individual items on the core form may be found in Appendix Q.

**Table 11.1. Reading Core Test: P-value Counts by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	0–9	0	0	0	0	0	0	0	0
	10–19	0	0	0	0	0	0	0	0
	20–29	0	0	0	0	0	0	1	0
	30–39	1	0	0	2	2	1	1	0
	40–49	4	7	6	4	4	2	7	1
	50–59	6	12	8	10	5	7	8	12
	60–69	10	8	10	7	13	16	8	13
	70–79	11	7	12	12	10	16	12	11
	80–89	7	6	6	7	9	2	8	8
	90–99	1	0	0	0	1	0	0	0
Summary	Total	40	40	42	42	44	44	45	45
	Mean	67.93	62.92	66.21	66.02	68.34	67.03	64.95	68.34
	Median	69.19	61.75	68.79	67.93	67.82	66.52	67.66	67.25
	Std. Dev.	14.05	13.64	12.27	13.61	14.69	10.93	16.24	11.35
	Minimum	31.04	41.24	43.21	35.38	37.56	38.31	25.63	44.33
	Maximum	91.96	89.90	85.00	86.99	90.17	82.94	89.59	89.23

**Table 11.2. Mathematics Core Test: P-value Counts by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	0–9	0	0	0	0	0	0	0	0
	10–19	0	0	0	0	0	0	0	0
	20–29	0	1	0	0	0	0	3	0
	30–39	0	1	1	2	2	2	3	1
	40–49	1	2	7	3	4	6	7	4
	50–59	2	2	11	10	8	11	9	10
	60–69	6	12	6	13	11	6	11	9
	70–79	14	9	7	10	15	12	2	10
	80–89	14	12	10	5	4	6	8	11
	90–99	3	1	3	2	1	2	2	0
Summary	Total	40	40	45	45	45	45	45	45
	Mean	76.28	70.78	67.03	66.45	66.02	64.84	60.29	67.88
	Median	76.09	73.18	65.36	68.03	68.67	68.47	60.70	67.91
	Std. Dev.	10.92	14.66	16.31	12.84	13.09	14.87	17.82	13.48
	Minimum	49.81	28.90	39.10	38.88	37.93	34.83	24.67	39.91
	Maximum	95.42	92.00	93.50	95.63	96.49	92.41	91.41	89.30

**Table 11.3. Language Usage Core Test: P-value Counts by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	0–9	0	0	0	0	0	0	0	0
	10–19	1	1	0	0	0	0	1	0
	20–29	0	3	2	0	0	0	1	1
	30–39	2	2	0	2	2	1	3	3
	40–49	1	5	9	4	8	4	8	7
	50–59	1	4	0	5	7	7	12	7
	60–69	6	6	5	10	9	9	5	11
	70–79	7	8	16	10	9	12	9	6
	80–89	13	3	8	7	8	7	4	8
	90–99	9	8	2	4	2	5	2	2
Summary	Total	40	40	42	42	45	45	45	45
	Mean	76.91	65.48	67.41	69.26	66.35	70.37	59.77	63.15
	Median	82.43	69.26	72.27	70.21	68.20	72.53	58.61	63.70
	Std. Dev.	18.79	23.14	18.30	16.18	15.16	14.77	18.37	17.78
	Minimum	14.07	15.40	20.46	31.24	31.96	35.66	13.27	26.82
	Maximum	96.43	97.59	94.54	97.32	93.83	96.97	94.62	96.94

**Table 11.4. Science Core Test: P-value Counts by Grade**

		<b>Grade</b>		
		<b>5</b>	<b>7</b>	<b>10</b>
<b>Range</b>	0–9	0	0	0
	10–19	0	0	0
	20–29	1	1	1
	30–39	1	2	3
	40–49	9	6	5
	50–59	10	12	12
	60–69	9	16	9
	70–79	4	6	16
	80–89	6	4	4
	90–99	2	1	0
<b>Summary</b>	Total	42	48	50
	Mean	61.83	61.09	61.97
	Median	60.25	61.03	63.99
	Std. Dev.	16.27	13.32	14.23
	Minimum	26.16	26.74	28.49
	Maximum	93.22	90.34	88.22

**Table 11.5. Reading Core Test: Item-Total Correlation Counts by Grade**

		<b>Grade</b>							
		<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>Range</b>	<0.0	0	0	0	0	0	0	0	0
	0.0–0.09	0	0	0	0	0	0	1	0
	0.10–0.19	0	0	0	0	0	1	3	0
	0.20–0.29	4	3	3	3	1	4	6	3
	0.30–0.39	4	10	7	14	10	12	14	9
	0.40–0.49	21	20	19	20	27	19	16	25
	0.50–0.59	11	7	13	5	6	8	4	8
	0.60–0.69	0	0	0	0	0	0	1	0
	0.70–0.79	0	0	0	0	0	0	0	0
	0.80–0.89	0	0	0	0	0	0	0	0
	0.90–0.99	0	0	0	0	0	0	0	0
<b>Summary</b>	Total	40	40	42	42	44	44	45	45
	Mean	0.45	0.43	0.45	0.42	0.44	0.41	0.38	0.43
	Median	0.45	0.44	0.46	0.43	0.44	0.42	0.39	0.44
	Std. Dev.	0.08	0.08	0.08	0.08	0.06	0.09	0.12	0.07
	Minimum	0.24	0.28	0.22	0.26	0.27	0.17	0.07	0.25
	Maximum	0.57	0.60	0.59	0.55	0.58	0.59	0.62	0.54

**Table 11.6. Mathematics Core Test: Item-Total Correlation Counts by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
		<0.0	0	0	0	0	0	0	0
	0.0–0.09	0	0	0	0	0	0	0	0
	0.10–0.19	1	0	0	0	0	2	1	0
	0.20–0.29	1	4	3	3	1	3	4	6
	0.30–0.39	9	11	16	15	10	7	8	13
	0.40–0.49	20	22	17	20	23	20	17	16
	0.50–0.59	9	3	9	6	9	13	15	9
	0.60–0.69	0	0	0	1	2	0	0	1
	0.70–0.79	0	0	0	0	0	0	0	0
	0.80–0.89	0	0	0	0	0	0	0	0
	0.90–0.99	0	0	0	0	0	0	0	0
Summary	Total	40	40	45	45	45	45	45	45
	Mean	0.44	0.41	0.43	0.42	0.45	0.43	0.44	0.42
	Median	0.44	0.43	0.44	0.43	0.46	0.45	0.45	0.42
	Std. Dev.	0.08	0.08	0.08	0.08	0.08	0.10	0.10	0.10
	Minimum	0.19	0.22	0.25	0.24	0.26	0.17	0.16	0.21
	Maximum	0.58	0.54	0.58	0.64	0.61	0.57	0.59	0.60

**Table 11.7. Language Usage Core Test: Item-Total Correlation Counts by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
		<0.0	0	0	0	1	0	0	0
	0.0–0.09	1	0	1	0	0	0	2	0
	0.10–0.19	1	1	0	0	1	0	4	3
	0.20–0.29	2	14	7	6	8	5	13	10
	0.30–0.39	11	13	17	20	10	22	13	17
	0.40–0.49	22	12	16	15	23	17	9	13
	0.50–0.59	3	0	1	1	2	1	4	2
	0.60–0.69	0	0	0	0	0	0	0	0
	0.70–0.79	0	0	0	0	0	0	0	0
	0.80–0.89	0	0	0	0	0	0	0	0
	0.90–0.99	0	0	0	0	0	0	0	0
Summary	Total	40	40	42	42	45	45	45	45
	Mean	0.41	0.34	0.37	0.38	0.37	0.37	0.33	0.35
	Median	0.42	0.35	0.38	0.38	0.40	0.37	0.33	0.35
	Std. Dev.	0.09	0.09	0.08	0.07	0.10	0.07	0.11	0.09
	Minimum	0.09	0.13	0.09	0.23	-0.04	0.23	0.07	0.15
	Maximum	0.57	0.48	0.52	0.53	0.56	0.50	0.52	0.52



**Table 11.8. Science Core Test: Item-Total Correlation Counts by Grade**

		Grade		
		5	7	10
Range	<0.0	0	0	0
	0.0–0.09	0	0	0
	0.10–0.19	1	0	2
	0.20–0.29	11	6	7
	0.30–0.39	18	19	9
	0.40–0.49	11	20	18
	0.50–0.59	1	3	14
	0.60–0.69	0	0	0
	0.70–0.79	0	0	0
	0.80–0.89	0	0	0
	0.90–0.99	0	0	0
Summary	Total	42	48	50
	Mean	0.34	0.39	0.41
	Median	0.35	0.39	0.45
	Std. Dev.	0.08	0.08	0.11
	Minimum	0.19	0.24	0.13
	Maximum	0.51	0.52	0.56

## 11.2 Rasch Summaries for the Core Form

Item calibrations for the core test were performed using Rasch analysis. As stated in section 6.4 as part of the field test analysis, a more useful comparison across grades may be made with item difficulties than with p-values given that the item difficulties for all grades within content are on the same scale. Tables 11.9–11.12 show the Rasch item difficulties in summary form across grades. These values are presented in logits, which are in the final scale metric save for a linear transformation. That is, multiplying the logits by 10 and adding 200 places the logit in the final scale score metric. Large negative logits represent easier items while large positive logits represent more difficult items.

The summary item difficulties are presented in groups of one half logit from less than -3.0, and thereafter in .5 logit blocks, with the final block at 6.0 logits and above. As you can see from the table of the difficulties for reading, most of the items fall into the -2.0 to 0.49 range for grade 3 and slowly increase in difficulty as grades increase to the right, an expected result. The last column on the right for grade 10 shows that most of the items fall into the 0.5 to 2.99 range. At the bottom of each table is the mean, median, standard deviation, minimum, and maximum for each grade. As can be expected, these descriptive values increase as the grades increase. The notable exception can be found in the reading table where the mean of the item difficulties for grade 10 is smaller than the mean for grade 9 (but not smaller than grade 8, as was the case in the embedded field test table (see section 6.4). Given that the grade 9 assessment in Spring 2007 consisted of items developed by the former vendor, the pattern of descriptive statistics (means and standard deviation) may not hold. Note that the one descriptive statistic that does not follow this progression is the standard deviation. The following tables for mathematics, language usage, and science follow the same pattern that is shown in reading. The exceptions: the mean for mathematics in grade 10 was lower than for grade 9 and the means for language usage were about the same for grades 7 and 8.

Rasch item parameters and fit statistics for individual items on the core form may be found in Appendix R.

**Table 11.9. Reading Core Test: Item Difficulty Summary by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	< -3.000	0	0	0	0	0	0	0	0
	-3.00 to -2.51	1	0	0	0	0	0	0	0
	-2.50 to -2.01	3	0	0	0	0	0	0	0
	-2.00 to -1.51	4	2	0	0	0	0	0	0
	-1.50 to -1.01	5	3	0	0	0	0	0	0
	-1.00 to -0.51	11	5	6	4	3	0	0	0
	-0.50 to -0.01	9	8	6	4	5	0	0	0
	0.00 to 0.49	5	10	14	10	7	0	2	2
	0.50 to 0.99	1	6	6	7	5	9	4	6
	1.00 to 1.49	1	6	7	10	13	9	11	8
	1.50 to 1.99	0	0	3	4	5	15	8	7
	2.00 to 2.49	0	0	0	3	3	8	4	16
	2.50 to 2.99	0	0	0	0	3	2	10	5
	3.00 to 3.49	0	0	0	0	0	1	4	1
	3.50 to 3.99	0	0	0	0	0	0	1	0
	4.00 to 4.49	0	0	0	0	0	0	1	0
	4.50 to 4.99	0	0	0	0	0	0	0	0
	5.00 to 5.49	0	0	0	0	0	0	0	0
	5.50 to 5.99	0	0	0	0	0	0	0	0
	>= 6.000	0	0	0	0	0	0	0	0
Summary	Total	40	40	42	42	44	44	45	45
	Mean	-0.744	0.039	0.409	0.716	0.958	1.621	1.963	1.801
	Median	-0.723	0.162	0.314	0.676	1.090	1.686	1.907	1.931
	Std. Dev.	0.889	0.798	0.717	0.780	0.926	0.611	0.907	0.705
	Minimum	-2.680	-1.855	-0.839	-0.676	-0.680	0.619	0.276	0.266
	Maximum	1.322	1.208	1.673	2.338	2.703	3.124	4.013	3.134

**Table 11.10. Mathematics Core Test: Item Difficulty Summary by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	< -3.000	1	0	0	0	0	0	0	0
	-3.00 to -2.51	2	0	0	0	0	0	0	0
	-2.50 to -2.01	4	0	0	0	0	0	0	0
	-2.00 to -1.51	6	2	0	0	0	0	0	0
	-1.50 to -1.01	11	3	3	1	0	0	0	0
	-1.00 to -0.51	10	10	2	1	1	0	0	0
	-0.50 to -0.01	3	6	7	0	0	0	0	0
	0.00 to 0.49	3	10	6	1	0	0	0	0
	0.50 to 0.99	0	5	5	7	1	2	0	0
	1.00 to 1.49	0	1	8	8	8	1	2	0
	1.50 to 1.99	0	2	7	12	14	8	1	2
	2.00 to 2.49	0	1	7	10	7	8	7	6
	2.50 to 2.99	0	0	0	3	8	7	2	8
	3.00 to 3.49	0	0	0	2	4	8	7	7
	3.50 to 3.99	0	0	0	0	2	7	11	10
	4.00 to 4.49	0	0	0	0	0	4	6	7
	4.50 to 4.99	0	0	0	0	0	0	4	5
	5.00 to 5.49	0	0	0	0	0	0	3	0
	5.50 to 5.99	0	0	0	0	0	0	2	0
	>= 6.000	0	0	0	0	0	0	0	0
Summary	Total	40	40	45	45	45	45	45	45
	Mean	-1.215	-0.110	0.763	1.581	2.065	2.734	3.625	3.389
	Median	-1.097	-0.164	0.985	1.566	1.974	2.603	3.676	3.461
	Std. Dev.	0.806	0.897	1.059	0.839	0.850	0.901	1.077	0.822
	Minimum	-3.278	-1.820	-1.416	-1.064	-0.933	0.600	1.361	1.841
	Maximum	0.426	2.226	2.388	3.101	3.662	4.410	5.686	4.940

**Table 11.11. Language Usage Core Test: Item Difficulty Summary by Grade**

		Grade							
Range		3	4	5	6	7	8	9	10
	< -3.000	7	1	0	0	0	0	0	0
	-3.00 to -2.51	7	1	0	0	0	0	0	0
	-2.50 to -2.01	5	4	0	1	0	0	0	0
	-2.00 to -1.51	5	2	1	1	0	1	0	0
	-1.50 to -1.01	4	1	3	1	1	1	0	1
	-1.00 to -0.51	7	5	2	3	1	0	1	0
	-0.50 to -0.01	1	5	14	3	1	4	1	0
	0.00 to 0.49	1	6	7	7	8	6	3	1
	0.50 to 0.99	0	4	4	10	7	3	3	8
	1.00 to 1.49	2	5	1	8	9	12	7	2
	1.50 to 1.99	0	2	8	3	8	9	5	6
	2.00 to 2.49	0	2	0	4	7	5	12	9
	2.50 to 2.99	1	1	2	1	3	3	8	8
	3.00 to 3.49	0	1	0	0	0	1	3	8
	3.50 to 3.99	0	0	0	0	0	0	1	1
	4.00 to 4.49	0	0	0	0	0	0	0	1
	4.50 to 4.99	0	0	0	0	0	0	1	0
	5.00 to 5.49	0	0	0	0	0	0	0	0
	5.50 to 5.99	0	0	0	0	0	0	0	0
	>= 6.000	0	0	0	0	0	0	0	0
Summary	Total	40	40	42	42	45	45	45	45
	Mean	-1.708	-0.010	0.346	0.678	1.181	1.174	1.983	2.079
	Median	-1.858	0.041	0.190	0.791	1.181	1.202	2.113	2.178
	Std. Dev.	1.412	1.537	1.065	1.093	0.888	0.998	1.043	1.064
	Minimum	-3.893	-3.114	-1.932	-2.259	-1.032	-1.561	-0.670	-1.017
	Maximum	2.612	3.005	2.966	2.805	2.992	3.067	4.684	4.002

**Table 11.12. Science Core Test: Item Difficulty Summary by Grade**

		Grade		
		5	7	10
Range	< -3.000	0	0	0
	-3.00 to -2.51	0	0	0
	-2.50 to -2.01	2	0	0
	-2.00 to -1.51	2	0	0
	-1.50 to -1.01	3	2	0
	-1.00 to -0.51	4	3	0
	-0.50 to -0.01	7	2	2
	0.00 to 0.49	10	10	4
	0.50 to 0.99	10	16	12
	1.00 to 1.49	3	12	11
	1.50 to 1.99	1	2	10
	2.00 to 2.49	0	1	7
	2.50 to 2.99	0	0	2
	3.00 to 3.49	0	0	2
	3.50 to 3.99	0	0	0
	4.00 to 4.49	0	0	0
	4.50 to 4.99	0	0	0
	5.00 to 5.49	0	0	0
	5.50 to 5.99	0	0	0
	>= 6.000	0	0	0
Summary	Total	42	48	50
	Mean	-0.015	0.600	1.393
	Median	0.157	0.657	1.331
	Std. Dev.	0.928	0.752	0.787
	Minimum	-2.306	-1.412	-0.344
	Maximum	1.841	2.407	3.173

### 11.3 Raw and Scale Score Descriptive Information for the Core Forms

Descriptive statistics are provided in Tables 11.13 to 11.20 for each content and grade. These statistics are presented for the raw score total, the scale score total, and each of the subscores in the raw score metric. The statistics include:

- number of students taking each assessment
- minimum
- maximum
- mean
- median
- standard deviation
- variance
- skewness, and
- kurtosis.

**Table 11.13. Grade 3 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19872	1	40	27.172	29	7.875	62.008	-0.663	-0.469
Scale Score	19872	152	245	203.022	204	11.910	141.842	-0.049	-0.079
Reading Process	19872	0	16	10.499	11	3.238	10.486	-0.513	-0.405
Comprehension/ Interpretation	19872	1	24	16.672	18	5.159	26.613	-0.704	-0.460

**Table 11.14. Grade 4 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19624	3	40	25.166	26	7.853	61.671	-0.271	-0.819
Scale Score	19624	173	252	207.825	208	11.421	130.449	0.341	0.200
Reading Process	19624	0	12	8.507	9	2.544	6.473	-0.630	-0.280
Comprehension/ Interpretation	19624	1	28	16.659	17	5.794	33.570	-0.158	-0.887

**Table 11.15. Grade 5 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19880	1	42	27.807	29	8.540	72.929	-0.573	-0.568
Scale Score	19880	165	256	213.284	213	11.895	141.499	0.051	-0.031
Reading Process	19880	0	11	7.147	8	2.548	6.491	-0.464	-0.583
Comprehension/ Interpretation	19880	0	31	20.660	22	6.436	41.425	-0.608	-0.509

**Table 11.16. Grade 6 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19421	1	42	27.730	29	7.825	61.237	-0.512	-0.470
Scale Score	19421	167	259	216.063	216	10.914	119.123	0.152	0.334
Reading Process	19421	0	12	8.471	9	2.542	6.463	-0.726	-0.053
Comprehension/ Interpretation	19421	0	30	19.260	20	5.767	33.254	-0.421	-0.582

**Table 11.17. Grade 7 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19683	0	44	30.071	31	8.404	70.622	-0.508	-0.513
Scale Score	19683	156	263	220.721	220	12.435	154.632	0.290	0.264
Reading Process	19683	0	12	9.413	10	2.386	5.695	-1.085	0.734
Comprehension/ Interpretation	19683	0	32	20.658	21	6.494	42.170	-0.332	-0.768

**Table 11.18. Grade 8 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19840	0	44	29.491	31	8.309	69.045	-0.583	-0.386
Scale Score	19840	165	268	225.307	226	10.842	117.539	0.089	0.343
Reading Process	19840	0	11	7.323	8	2.450	6.000	-0.500	-0.477
Comprehension/ Interpretation	19840	0	33	22.168	23	6.338	40.170	-0.618	-0.305

**Table 11.19. Grade 9 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	20544	0	44	29.225	31	7.491	56.111	-0.800	0.257
Scale Score	20544	166	261	227.518	229	9.295	86.392	-0.583	1.017
Reading Process	20544	0	9	6.716	7	1.876	3.519	-1.108	1.091
Comprehension/ Interpretation	20544	0	35	22.510	24	6.072	36.868	-0.657	-0.031

**Table 11.20. Grade 10 Descriptive Statistics for Reading**

<b>Reading</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19092	0	45	30.752	32	8.741	76.405	-0.615	-0.332
Scale Score	19092	165	270	228.585	228	11.782	138.807	0.148	0.616
Reading Process	19092	0	11	7.765	8	2.379	5.658	-0.735	-0.043
Comprehension/ Interpretation	19092	0	34	22.986	24	6.795	46.169	-0.562	-0.425

**Table 11.21. Grade 3 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19915	5	40	30.513	32	7.135	50.905	-0.882	0.020
Scale Score	19915	166	240	204.338	203	13.206	174.393	0.272	-0.026
Number and Operation	19915	0	15	12.105	13	2.983	8.900	-1.151	0.672
Concepts and Principles of Measurement	19915	0	7	4.809	5	1.576	2.485	-0.552	-0.304
Concepts and Language of Algebra and Functions	19915	0	6	4.225	4	1.477	2.182	-0.630	-0.344
Concepts and Principles of Geometry	19915	0	6	5.043	5	1.166	1.358	-1.286	1.279
Data Analysis, Probability, and Statistics	19915	0	6	4.331	5	1.595	2.545	-0.805	-0.280



**Table 11.22. Grade 4 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19680	4	40	28.314	30	7.024	49.343	-0.694	-0.181
Scale Score	19680	174	252	211.099	212	11.481	131.815	0.256	0.450
Number and Operation	19680	0	15	11.700	13	3.104	9.637	-1.036	0.409
Concepts and Principles of Measurement	19680	0	6	4.061	4	1.424	2.029	-0.568	-0.292
Concepts and Language of Algebra and Functions	19680	0	7	4.653	5	1.552	2.409	-0.493	-0.269
Concepts and Principles of Geometry	19680	0	6	3.935	4	1.431	2.046	-0.336	-0.574
Data Analysis, Probability, and Statistics	19680	0	6	3.965	4	1.510	2.280	-0.513	-0.475

**Table 11.23. Grade 5 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19928	4	45	30.164	31	8.552	73.133	-0.371	-0.692
Scale Score	19928	180	262	218.395	218	12.618	159.203	0.417	0.325
Number and Operation	19928	0	15	10.037	10	3.318	11.007	-0.370	-0.708
Concepts and Principles of Measurement	19928	0	7	4.205	4	1.745	3.044	-0.144	-0.867
Concepts and Language of Algebra and Functions	19928	0	8	5.448	6	1.785	3.185	-0.413	-0.549
Concepts and Principles of Geometry	19928	0	8	5.581	6	1.806	3.262	-0.593	-0.313
Data Analysis, Probability, and Statistics	19928	0	7	4.894	5	1.650	2.724	-0.645	-0.212

**Table 11.24. Grade 6 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19477	2	45	29.902	31	8.661	75.017	-0.298	-0.797
Scale Score	19477	181	269	225.782	225	12.073	145.758	0.589	0.623
Number and Operation	19477	0	12	7.771	8	2.759	7.615	-0.238	-0.859
Concepts and Principles of Measurement	19477	0	7	4.843	5	1.685	2.838	-0.524	-0.519
Concepts and Language of Algebra and Functions	19477	0	10	6.576	7	2.308	5.325	-0.359	-0.713
Concepts and Principles of Geometry	19477	0	9	5.778	6	2.028	4.114	-0.312	-0.591
Data Analysis, Probability, and Statistics	19477	0	7	4.934	5	1.673	2.800	-0.574	-0.434

**Table 11.25. Grade 7 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19714	3	45	29.711	31	9.302	86.535	-0.360	-0.836
Scale Score	19714	191	274	230.460	230	13.194	174.087	0.483	0.219
Number and Operation	19714	0	12	8.350	9	3.173	10.065	-0.669	-0.640
Concepts and Principles of Measurement	19714	0	7	4.367	5	1.814	3.290	-0.336	-0.752
Concepts and Language of Algebra and Functions	19714	0	9	5.649	6	2.194	4.813	-0.271	-0.802
Concepts and Principles of Geometry	19714	0	9	5.435	5	2.099	4.408	-0.153	-0.706
Data Analysis, Probability, and Statistics	19714	0	8	5.909	6	1.745	3.044	-0.710	-0.156

**Table 11.26. Grade 8 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19864	2	45	29.178	30	8.829	77.943	-0.282	-0.884
Scale Score	19864	193	281	236.302	236	12.151	147.653	0.438	0.192
Number and Operation	19864	0	11	7.270	8	2.571	6.608	-0.375	-0.770
Concepts and Principles of Measurement	19864	0	6	3.193	3	1.459	2.128	0.039	-0.713
Concepts and Language of Algebra and Functions	19864	0	13	8.795	9	3.185	10.143	-0.444	-0.843
Concepts and Principles of Geometry	19864	0	9	5.810	6	1.930	3.726	-0.308	-0.590
Data Analysis, Probability, and Statistics	19864	0	6	4.110	4	1.486	2.207	-0.536	-0.451

**Table 11.27. Grade 9 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	20611	0	45	27.131	28	9.027	81.493	-0.235	-0.795
Scale Score	20611	181	291	242.629	243	12.303	151.355	0.271	0.223
Number and Operation	20611	0	9	5.929	6	1.946	3.788	-0.458	-0.430
Concepts and Principles of Measurement	20611	0	9	5.136	5	2.391	5.716	-0.095	-0.946
Concepts and Language of Algebra and Functions	20611	0	13	8.454	9	3.045	9.271	-0.501	-0.601
Concepts and Principles of Geometry	20611	0	7	3.409	3	1.702	2.898	0.191	-0.655
Data Analysis, Probability, and Statistics	20611	0	7	4.203	4	1.656	2.742	-0.293	-0.528

**Table 11.28. Grade 10 Descriptive Statistics for Mathematics**

<b>Mathematics</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19131	0	45	30.546	32	8.417	70.839	-0.484	-0.476
Scale Score	19131	181	287	244.339	244	11.869	140.881	0.365	0.565
Number and Operation	19131	0	7	4.586	5	1.805	3.259	-0.441	-0.683
Concepts and Principles of Measurement	19131	0	7	4.592	5	1.618	2.619	-0.518	-0.325
Concepts and Language of Algebra and Functions	19131	0	14	9.828	10	3.063	9.383	-0.528	-0.485
Concepts and Principles of Geometry	19131	0	9	5.906	6	2.045	4.182	-0.494	-0.426
Data Analysis, Probability, and Statistics	19131	0	8	5.634	6	1.616	2.612	-0.533	-0.122

**Table 11.29. Grade 3 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19878	5	40	30.763	32	6.004	36.042	-1.078	0.874
Scale Score	19878	158	245	201.227	201	12.570	158.015	0.032	0.217
Writing Process	19878	1	16	12.842	13	2.671	7.134	-1.142	1.214
Writing Components	19878	2	24	17.921	19	3.800	14.440	-1.002	0.640

**Table 11.30. Grade 4 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19631	1	40	26.191	27	5.572	31.047	-0.417	-0.094
Scale Score	19631	153	258	210.051	211	9.286	86.230	0.033	0.729
Writing Process	19631	0	17	11.258	11	2.643	6.983	-0.498	0.106
Writing Components	19631	1	23	14.933	15	3.508	12.304	-0.340	-0.179

**Table 11.31. Grade 5 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19887	4	42	28.312	29	6.635	44.024	-0.540	-0.179
Scale Score	19887	177	259	213.330	213	9.943	98.858	0.145	0.255
Writing Process	19887	1	22	15.518	16	3.742	14.003	-0.690	0.055
Writing Components	19887	0	20	12.794	13	3.466	12.014	-0.347	-0.367

**Table 11.32. Grade 6 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19431	4	42	29.090	30	6.903	47.654	-0.577	-0.204
Scale Score	19431	180	261	218.396	218	10.511	110.487	0.188	0.449
Writing Process	19431	0	21	13.944	15	4.105	16.848	-0.523	-0.446
Writing Components	19431	1	21	15.146	16	3.375	11.388	-0.634	0.203

**Table 11.33. Grade 7 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19682	0	45	29.858	31	7.412	54.938	-0.479	-0.356
Scale Score	19682	158	265	220.822	221	9.542	91.040	0.062	0.237
Writing Process	19682	0	26	17.876	19	4.791	22.955	-0.616	-0.255
Writing Components	19682	0	19	11.982	12	3.246	10.535	-0.258	-0.404

**Table 11.34. Grade 8 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19821	2	45	31.668	33	7.140	50.984	-0.684	0.021
Scale Score	19821	176	266	223.618	224	9.912	98.244	0.087	0.535
Writing Process	19821	0	26	18.534	19	4.717	22.250	-0.745	-0.019
Writing Components	19821	1	19	13.134	13	3.025	9.150	-0.473	-0.027

**Table 11.35. Grade 9 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	20523	0	45	26.896	27	6.783	46.014	-0.172	-0.432
Scale Score	20523	164	275	225.169	225	8.359	69.870	0.200	0.686
Writing Process	20523	0	29	18.335	19	4.551	20.707	-0.340	-0.303
Writing Components	20523	0	16	8.560	9	2.877	8.276	0.025	-0.517

**Table 11.36. Grade 10 Descriptive Statistics for Language Usage**

<b>Language Usage</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19101	0	45	28.416	29	6.921	47.904	-0.414	-0.266
Scale Score	19101	164	275	228.353	228	8.754	76.640	-0.016	0.520
Writing Process	19101	0	28	18.876	20	4.635	21.486	-0.677	-0.028
Writing Components	19101	0	17	9.540	9	2.957	8.743	0.055	-0.370

**Table 11.37. Grade 5 Descriptive Statistics for Science**

<b>Science</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19840	2	42	25.969	26	6.581	43.309	-0.152	-0.485
Scale Score	19840	166	253	206.463	206	8.762	76.777	0.383	0.787
Nature of Science	19840	0	16	11.154	11	2.848	8.113	-0.496	-0.180
Physical Science	19840	0	6	3.379	3	1.454	2.113	0.051	-0.712
Biology	19840	0	7	3.708	4	1.601	2.564	0.096	-0.602
Earth and Space Systems	19840	0	6	2.785	3	1.350	1.822	0.154	-0.443
Personal and Social Perspectives; Technology	19840	0	7	4.942	5	1.582	2.501	-0.688	-0.072

**Table 11.38. Grade 7 Descriptive Statistics for Science**

<b>Science</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	19587	1	48	29.323	30	8.755	76.659	-0.201	-0.788
Scale Score	19587	165	259	211.941	212	9.924	98.492	0.273	0.165
Nature of Science	19587	0	21	13.650	14	4.289	18.394	-0.332	-0.689
Physical Science	19587	0	7	4.295	4	1.634	2.669	-0.247	-0.604
Biology	19587	0	8	4.926	5	1.770	3.135	-0.320	-0.476
Earth and Space Systems	19587	0	6	2.922	3	1.502	2.256	0.036	-0.693
Personal and Social Perspectives; Technology	19587	0	6	3.530	4	1.599	2.557	-0.180	-0.857

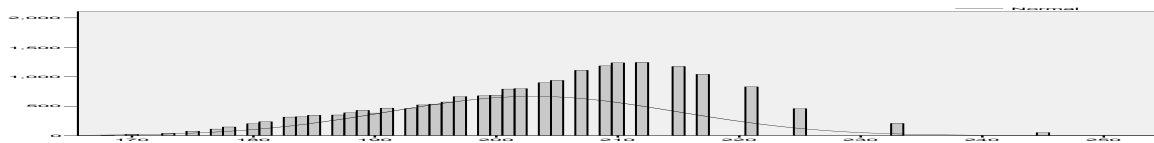
**Table 11.39. Grade 10 Descriptive Statistics for Science**

<b>Science</b>	<b>N</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Median</b>	<b>Std. Deviation</b>	<b>Variance</b>	<b>Skewness</b>	<b>Kurtosis</b>
Raw Score	18769	1	50	30.985	32	9.566	91.504	-0.272	-0.830
Scale Score	18769	172	268	220.461	220	10.797	116.581	0.271	0.159
Nature of Science	18769	0	21	12.971	13	4.463	19.915	-0.242	-0.836
Physical Science	18769	0	7	3.375	3	1.554	2.414	0.157	-0.425
Biology	18769	0	7	4.825	5	1.579	2.492	-0.507	-0.377
Earth and Space Systems	18769	0	7	4.630	5	1.684	2.835	-0.536	-0.435
Personal and Social Perspectives; Technology	18769	0	8	5.185	6	2.123	4.508	-0.527	-0.652

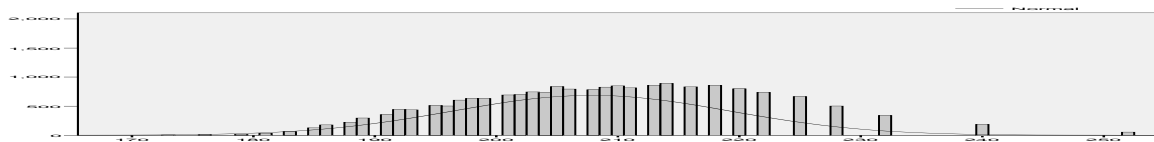
## 11.4 Student Scale Score Distributions Across Grades

Figures 11.1–11.4 present the student scale score distributions across grades for each content. As expected, the origins or centers of each distribution move progressively from left to right as they move from grade 3 to grade 10. These “centers” represent the approximate location of the mean or median of the performance of each grade. The figures also reveal the degree of overlap across grades. For example, in reading, one can see that high performing students in 4th grade are in about the same location as the average 8th grade students. A caution to the reader is made here given that the experiences in terms of development and instruction in content varies greatly from grade to grade and student to student. Thus, comparisons such as the above should be treated with caution, in particular when non-adjacent grades are involved.

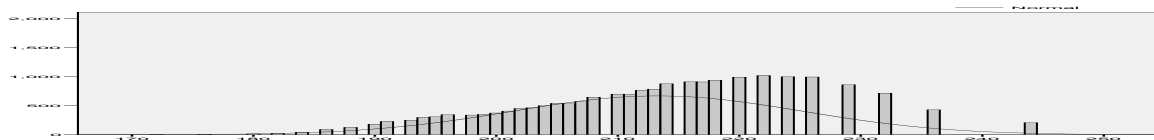
**Figure 11.1**  
**Grade 3 Reading**



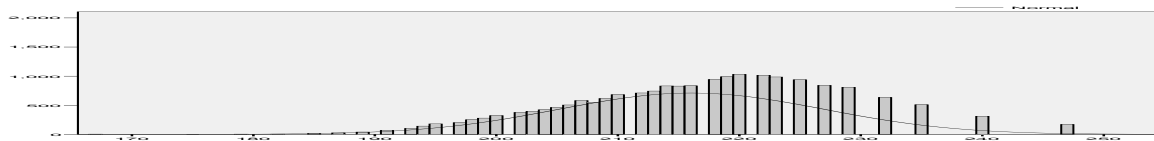
**Grade 4 Reading**



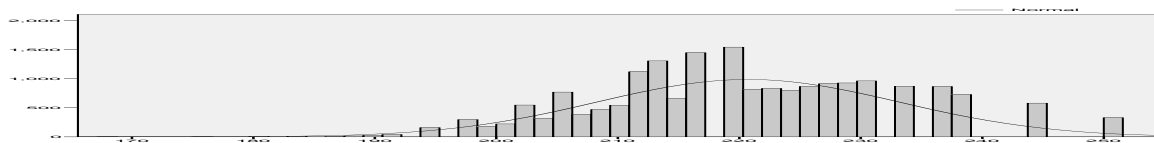
**Grade 5 Reading**



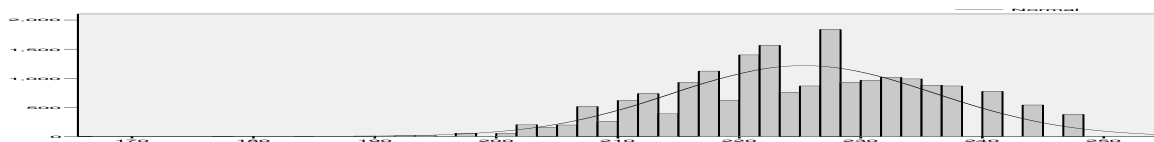
**Grade 6 Reading**



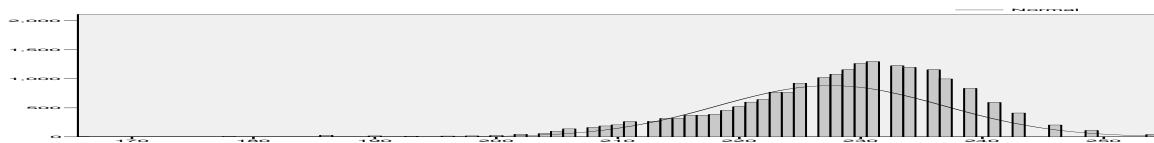
**Grade 7 Reading**



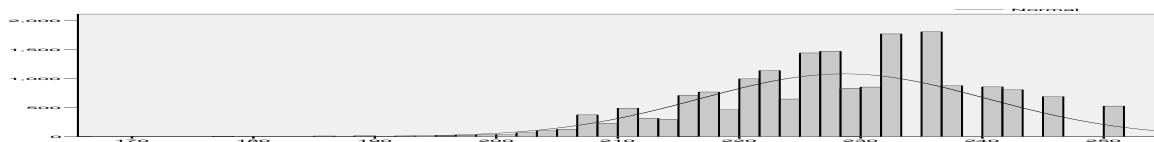
**Grade 8 Reading**



**Grade 9 Reading**

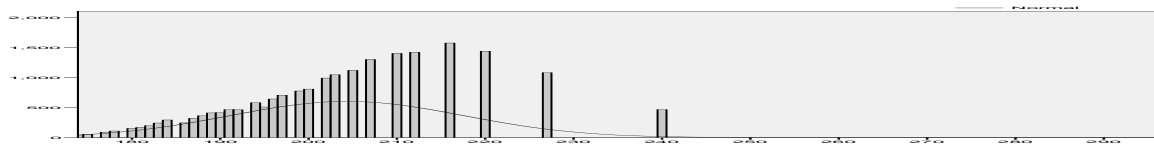


**Grade 10 Reading**

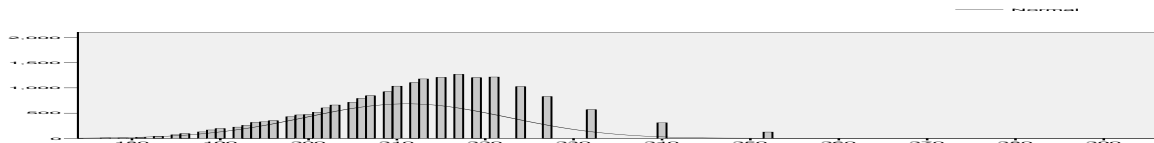




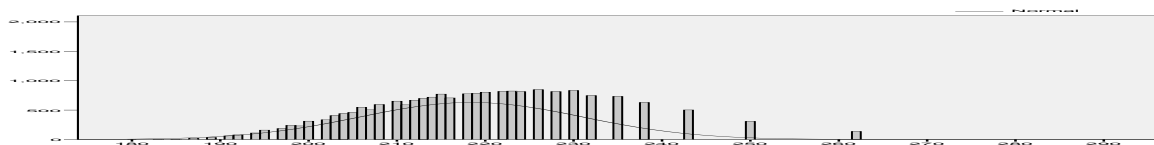
**Figure 11.2**  
**Grade 3 Mathematics**



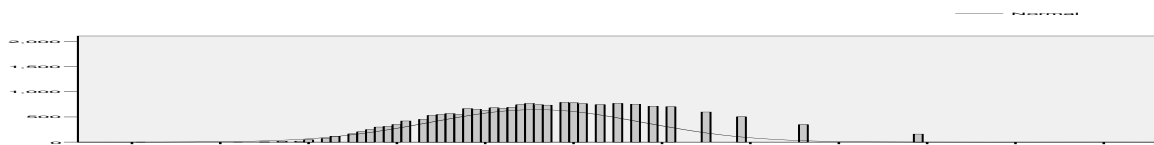
**Grade 4 Mathematics**



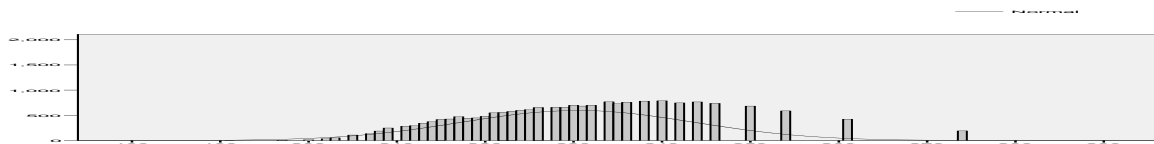
**Grade 5 Mathematics**



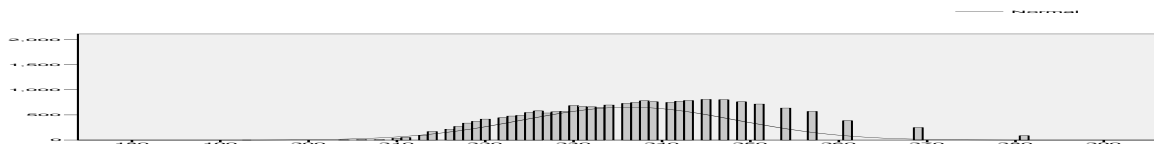
**Grade 6 Mathematics**



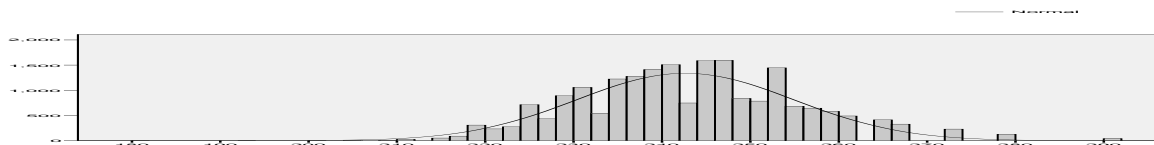
**Grade 7 Mathematics**



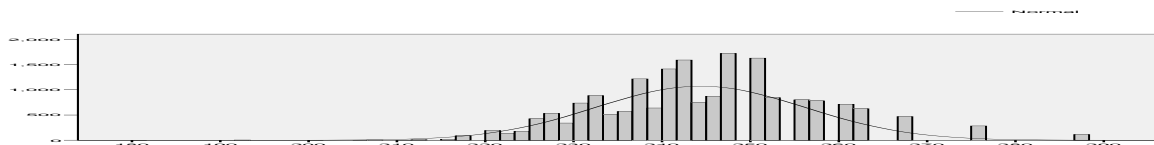
**Grade 8 Mathematics**



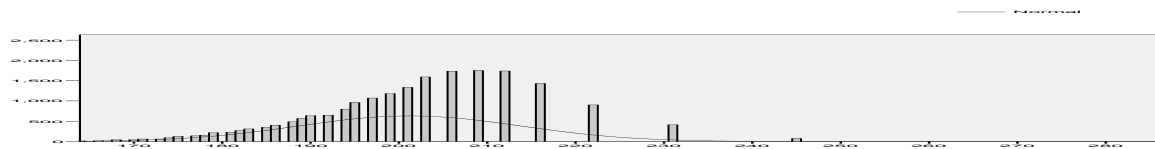
**Grade 9 Mathematics**



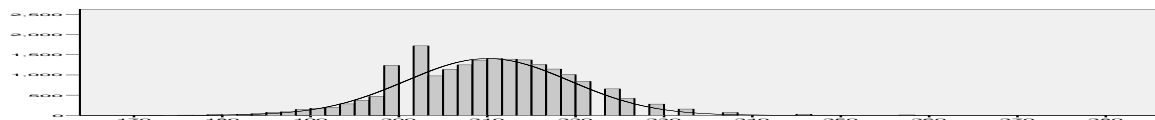
**Grade 10 Mathematics**



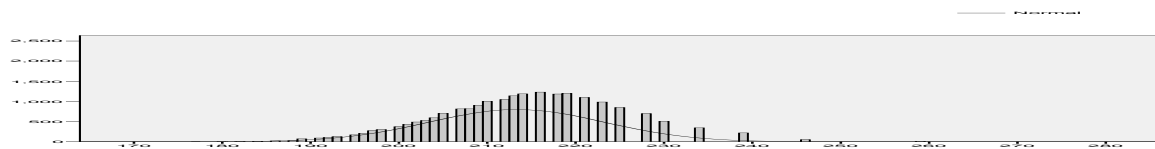
**Figure 11.3**  
**Grade 3 Language Usage**



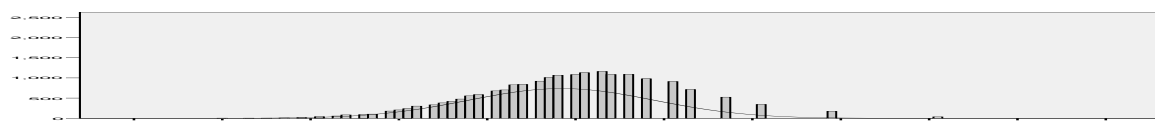
**Grade 4 Language Usage**



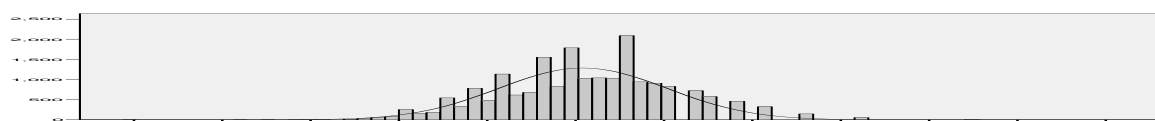
**Grade 5 Language Usage**



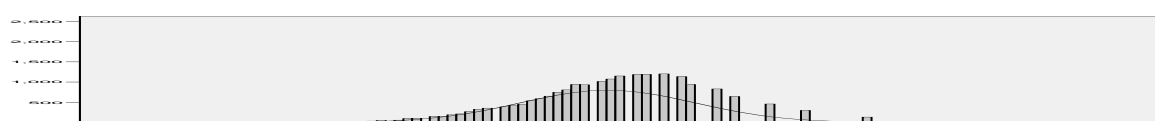
**Grade 6 Language Usage**



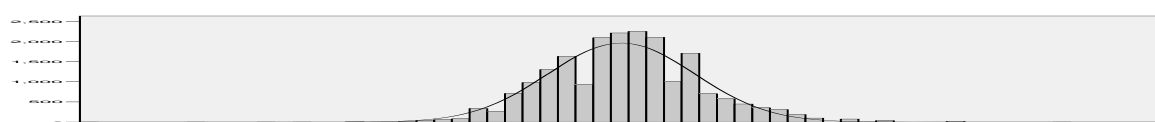
**Grade 7 Language Usage**



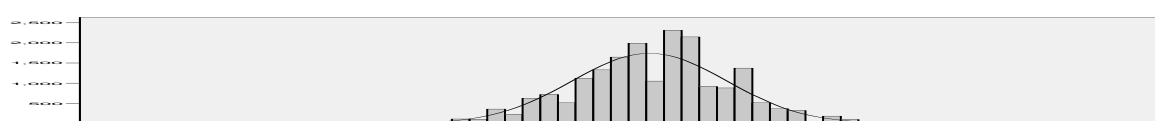
**Grade 8 Language Usage**



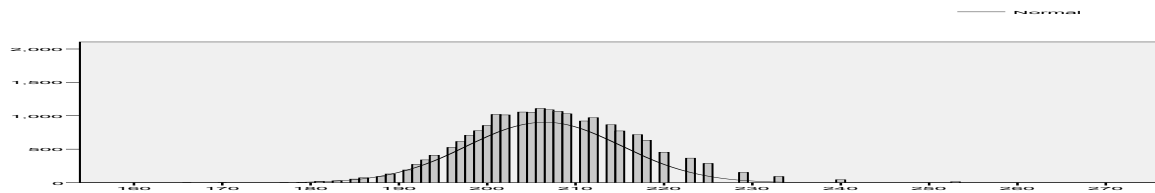
**Grade 9 Language Usage**



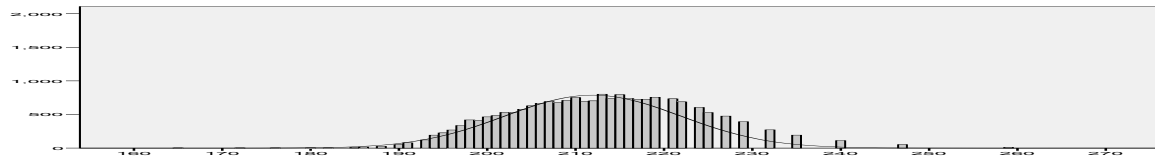
**Grade 10 Language Usage**



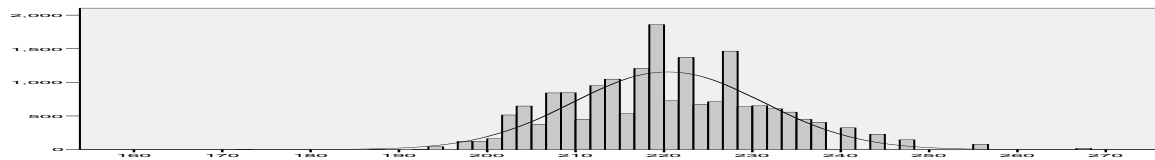
**Figure 11.4**  
**Grade 5 Science**



**Grade 7 Science**



**Grade 10 Science**



## 12. Reliability

The classical view of measurement considers all measures as having a “true” component and an error component. Errors occur as a natural part of the measurement process and can never be eliminated entirely. For example, uncontrollable factors such as differences in the testing environment and examinee disposition may increase error and decrease reliability. This is the fundamental premise of true-score reliability analysis and measurement theory. Stated explicitly, this relationship can be seen as the following:

$$X = T + E,$$

where  $X$  represents the observed test score,  $T$ , the student’s true score, and  $E$ , random error.

If the variance of the observed measures is denoted by  $\sigma_X^2$  and the variance of error by  $\sigma_E^2$  then the reliability ( $\rho_{XX}$ ) is given by:

$$\rho_{XX} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2}.$$

When there is no error, the reliability index is the true score variance divided by the true score variance, which is one. The variance of the observed measures can be estimated from the variance of the raw scores using the usual variance formula and the error variance can be estimated by:

$$\Sigma p(1-p),$$

where  $p$  is the proportion correct for each item.

The reliability index used for the 2007 administration of the ISAT was the Coefficient Alpha (Cronbach, 1951):

$$\alpha = \left( \frac{k}{k-1} \right) \left( 1 - \frac{\sum \sigma_i^2}{\sigma_X^2} \right),$$

where  $k$  is the number of items,  $\sigma_i^2$  is the variance of the set of scores associated with item  $i$ , and  $\sigma_X^2$  is the variance of the set of observed total scores.

Acceptable  $\alpha$  values generally range in the mid to high 0.80s to low 0.90s. When there is no error, the reliability index is the true score variance divided by the true score variance, which is one. Table 12.1 provides the total test Coefficient Alpha for each grade and academic content area combination. As can be seen in the tables, reading and mathematics fall into the acceptable range. Language usage in grades 4, 9, and 10 are somewhat lower, as is grade 5 for science. Overall, these  $\alpha$  values provide evidence of good reliability.

**Table 12.1. Total Test Reliabilities by Grade**

Grade	Reading		Mathematics		Lang. Usage		Science	
	N	Alpha	N	Alpha	N	Alpha	N	Alpha
3	19872	0.893	19915	0.889	19878	0.863	19840	0.815
4	19624	0.882	19680	0.871	19631	0.796		
5	19880	0.900	19928	0.900	19887	0.842		
6	19421	0.878	19477	0.896	19431	0.855		
7	19683	0.897	19714	0.912	19682	0.854	19587	0.880
8	19840	0.884	19864	0.901	19821	0.852		
9	20544	0.839	20611	0.904	20523	0.814		
10	19092	0.895	19131	0.891	19101	0.827	18769	0.899

Tables 12.2–12.5 contain the Coefficient Alphas for the academic content area standards.

**Table 12.2. Reading Content Standard Reliabilities by Grade**

		Reading Process		Comprehension/Interpretation	
Grade	Number of students	Number of items	Alpha	Number of items	Alpha
3	19872	16	0.745	24	0.854
4	19624	12	0.700	28	0.842
5	19880	11	0.697	31	0.872
6	19421	12	0.703	30	0.834
7	19683	12	0.721	32	0.867
8	19840	11	0.675	33	0.850
9	20544	9	0.540	36	0.803
10	19092	11	0.680	34	0.867

**Table 12.3. Mathematics Content Standard Reliabilities by Grade**

		Number and Operation		Concepts and Principles of Measurement		Concepts and Language of Algebra and Functions		Concepts and Principles of Geometry		Data Analysis, Probability, and Statistics	
Grade	Number of Students	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha
3	19915	15	0.798	7	0.505	6	0.529	6	0.530	6	0.675
4	19680	15	0.795	6	0.518	7	0.537	6	0.465	6	0.546
5	19928	15	0.786	7	0.607	8	0.612	8	0.591	7	0.596
6	19477	12	0.742	7	0.571	10	0.687	9	0.582	7	0.576
7	19714	12	0.822	7	0.612	9	0.671	9	0.624	8	0.619
8	19864	11	0.718	6	0.418	13	0.800	9	0.595	6	0.540
9	20611	9	0.612	9	0.720	13	0.775	7	0.546	7	0.557
10	19131	7	0.615	7	0.556	14	0.761	9	0.634	8	0.460

**Table 12.4. Language Usage Content Standard Reliabilities by Grade**

		Writing Process		Writing Components	
Grade	Number of students	Number of items	Alpha	Number of items	Alpha
3	19878	16	0.734	24	0.790
4	19631	17	0.610	23	0.707
5	19887	22	0.744	20	0.720
6	19431	21	0.781	21	0.721
7	19682	26	0.812	19	0.651
8	19821	26	0.805	19	0.658
9	20523	29	0.740	16	0.618
10	19101	28	0.770	17	0.624

**Table 12.5. Science Content Standard Reliabilities by Grade**

		Nature of Science		Physical Science		Biology		Earth and Space Systems		Personal and Social Perspectives; Technology	
Grade	Number of Students	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha	Number of Items	Alpha
5	19840	16	0.651	6	0.456	7	0.434	6	0.261	7	0.548
7	19587	21	0.797	7	0.478	8	0.504	6	0.455	6	0.527
10	18769	21	0.808	7	0.400	7	0.510	7	0.552	8	0.697

Tables 12.6–12.13 contain the Coefficient Alphas for each of the NCLB student subgroups.

**Table 12.6. Grade 3: Reliability**

	Reading		Mathematics		Lang. Usage	
	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19872	0.893	19915	0.889	19878	0.863
<b>GENDER</b>						
Male	10251	0.897	10274	0.892	10253	0.866
Female	9621	0.887	9641	0.887	9625	0.849
<b>ETHNICITY</b>						
American Indian/Alaskan Native	292	0.882	291	0.885	292	0.862
Asian	233	0.869	234	0.888	234	0.860
Black/African American	246	0.900	251	0.907	247	0.888
Hawaiian/Other Pacific Islander	90	0.889	91	0.893	90	0.843
White	15716	0.888	15721	0.884	15726	0.856
Hispanic or Latino	2940	0.870	2973	0.877	2935	0.859
Other/Unknown	355	0.888	354	0.882	354	0.868
<b>FREE OR REDUCED LUNCH</b>						
No	10659	0.878	10669	0.877	10657	0.846
Yes	9213	0.890	9246	0.887	9221	0.862
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	18077	0.888	18075	0.884	18085	0.857
LEP	1395	0.835	1398	0.863	1394	0.850
LEP in first year of school	29	0.834	71	0.890	29	0.916
LEP exited in past 1 year	240	0.843	240	0.873	240	0.815
LEP exited in past 2 years	131	0.818	131	0.839	130	0.825
<b>TITLE I</b>						
No	12854	0.885	12865	0.883	12856	0.853
Yes	7018	0.888	7050	0.884	7022	0.859
<b>MIGRANT STATUS</b>						
No	19523	0.892	19560	0.888	19529	0.861
Yes	349	0.858	355	0.873	349	0.883
<b>GIFTED AND TALENTED</b>						
No	19071	0.891	19115	0.886	19077	0.860
Yes	801	0.697	800	0.727	801	0.640
<b>NEGLECTED OR DELINQUENT</b>						
No	19872	0.893	19915	0.889	19878	0.863
Yes	0	N/A	0	N/A	0	N/A
<b>HOMELESS</b>						
No	19805	0.893	19848	0.889	19812	0.862
Yes	67	0.903	67	0.886	66	0.888
<b>HOME SCHOOL</b>						
No	19872	0.893	19915	0.889	19878	0.863
Yes	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>						
No	17608	0.881	17646	0.880	17609	0.849
Yes	2030	0.889	2034	0.896	2035	0.857
SPE exited in past 1 year	170	0.889	170	0.888	170	0.855
SPE exited in past 2 years	64	0.892	65	0.853	64	0.836
<b>504 PLAN</b>						
No	19730	0.893	19772	0.889	19735	0.863
Yes	142	0.877	143	0.880	143	0.846

**Table 12.7. Grade 4: Reliability**

	Reading		Mathematics		Lang. Usage	
	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19624	0.882	19680	0.871	19631	0.796
<b>GENDER</b>						
Male	10072	0.882	10102	0.872	10077	0.797
Female	9552	0.881	9578	0.870	9554	0.786
<b>ETHNICITY</b>						
American Indian/Alaskan Native	298	0.852	298	0.873	300	0.777
Asian	234	0.881	239	0.875	234	0.799
Black/African American	217	0.874	220	0.885	217	0.811
Hawaiian/Other Pacific Islander	72	0.855	72	0.878	72	0.734
White	15651	0.878	15662	0.863	15660	0.792
Hispanic or Latino	2782	0.843	2819	0.862	2780	0.761
Other/Unknown	370	0.886	370	0.876	368	0.785
<b>FREE OR REDUCED LUNCH</b>						
No	10740	0.873	10761	0.856	10746	0.783
Yes	8884	0.871	8919	0.870	8885	0.783
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	17836	0.878	17850	0.865	17847	0.792
LEP	1431	0.785	1432	0.844	1426	0.724
LEP in first year of school	16	0.630	57	0.847	17	0.790
LEP exited in past 1 year	225	0.809	225	0.856	225	0.670
LEP exited in past 2 years	116	0.835	116	0.846	116	0.732
<b>TITLE I</b>						
No	13236	0.877	13260	0.862	13242	0.788
Yes	6388	0.877	6420	0.875	6389	0.789
<b>MIGRANT STATUS</b>						
No	19317	0.881	19362	0.870	19327	0.795
Yes	307	0.822	318	0.865	304	0.767
<b>GIFTED AND TALENTED</b>						
No	18397	0.873	18453	0.866	18403	0.782
Yes	1227	0.782	1227	0.654	1228	0.649
<b>NEGLECTED OR DELINQUENT</b>						
No	19624	0.882	19680	0.871	19631	0.796
Yes	0	N/A	0	N/A	0	N/A
<b>HOMELESS</b>						
No	19580	0.882	19636	0.871	19587	0.795
Yes	44	0.773	44	0.838	44	0.755
<b>HOME SCHOOL</b>						
No	19624	0.882	19680	0.871	19631	0.796
Yes	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>						
No	17398	0.871	17451	0.859	17404	0.777
Yes	1997	0.857	2000	0.866	1998	0.762
SPE exited in past 1 year	150	0.885	150	0.878	150	0.784
SPE exited in past 2 years	79	0.885	79	0.854	79	0.797
<b>504 PLAN</b>						
No	19448	0.882	19504	0.871	19455	0.796
Yes	176	0.855	176	0.869	176	0.748



**Table 12.8. Grade 5: Reliability**

	Reading		Mathematics		Lang. Usage		Science	
	N	Alpha	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19880	0.900	19928	0.900	19887	0.842	19840	0.815
<b>GENDER</b>								
Male	10191	0.905	10213	0.906	10197	0.842	10164	0.831
Female	9689	0.895	9715	0.893	9690	0.836	9676	0.794
<b>ETHNICITY</b>								
American Indian/Alaskan Native	320	0.889	321	0.876	319	0.821	320	0.767
Asian	245	0.903	249	0.905	246	0.877	246	0.829
Black/African American	214	0.914	219	0.911	214	0.861	213	0.796
Hawaiian/Other Pacific Islander	81	0.887	81	0.908	81	0.832	81	0.853
White	15947	0.893	15953	0.896	15947	0.832	15908	0.809
Hispanic or Latino	2719	0.884	2749	0.883	2725	0.821	2721	0.764
Other/Unknown	354	0.906	356	0.901	355	0.836	351	0.831
<b>FREE OR REDUCED LUNCH</b>								
No	11137	0.888	11152	0.892	11142	0.823	11121	0.803
Yes	8743	0.896	8776	0.894	8745	0.834	8719	0.801
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18277	0.895	18280	0.897	18281	0.835	18236	0.810
LEP	1256	0.840	1257	0.850	1256	0.784	1254	0.706
LEP in first year of school	28	0.882	72	0.882	30	0.810	32	0.743
LEP exited in past 1 year	185	0.851	185	0.867	185	0.736	185	0.703
LEP exited in past 2 years	134	0.861	134	0.843	135	0.784	133	0.652
<b>TITLE I</b>								
No	14271	0.894	14294	0.896	14277	0.835	14255	0.807
Yes	5609	0.903	5634	0.899	5610	0.838	5585	0.815
<b>MIGRANT STATUS</b>								
No	19569	0.899	19615	0.899	19576	0.840	19530	0.814
Yes	311	0.884	313	0.893	311	0.827	310	0.758
<b>GIFTED AND TALENTED</b>								
No	18512	0.895	18560	0.892	18520	0.830	18474	0.799
Yes	1368	0.701	1368	0.773	1367	0.630	1366	0.707
<b>NEGLECTED OR DELINQUENT</b>								
No	19878	0.900	19926	0.900	19885	0.842	19838	0.815
Yes	2	N/A	2	N/A	2	N/A	2	N/A
<b>HOMELESS</b>								
No	19845	0.900	19893	0.900	19853	0.841	19806	0.815
Yes	35	0.868	35	0.888	34	0.830	34	0.812
<b>HOME SCHOOL</b>								
No	19880	0.900	19928	0.900	19887	0.842	19840	0.815
Yes	0	N/A	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>								
No	17639	0.886	17684	0.890	17643	0.823	17603	0.805
Yes	1987	0.887	1991	0.879	1991	0.809	1985	0.794
SPE exited in past 1 year	183	0.919	182	0.899	182	0.853	181	0.854
SPE exited in past 2 years	71	0.905	71	0.911	71	0.838	71	0.803
<b>504 PLAN</b>								
No	19660	0.900	19709	0.900	19668	0.842	19622	0.815
Yes	220	0.896	219	0.893	219	0.812	218	0.827

**Table 12.9. Grade 6: Reliability**

	Reading		Mathematics		Lang. Usage	
	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19421	0.878	19477	0.896	19431	0.855
<b>GENDER</b>						
Male	9858	0.883	9890	0.903	9863	0.860
Female	9563	0.871	9587	0.887	9568	0.840
<b>ETHNICITY</b>						
American Indian/Alaskan Native	309	0.863	310	0.881	309	0.851
Asian	233	0.891	237	0.908	233	0.870
Black/African American	216	0.882	219	0.898	217	0.869
Hawaiian/Other Pacific Islander	81	0.887	81	0.897	81	0.878
White	15589	0.871	15607	0.891	15596	0.845
Hispanic or Latino	2662	0.852	2691	0.867	2662	0.841
Other/Unknown	331	0.860	332	0.874	333	0.832
<b>FREE OR REDUCED LUNCH</b>						
No	11126	0.865	11145	0.890	11130	0.837
Yes	8295	0.873	8332	0.886	8301	0.850
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	17929	0.872	17943	0.892	17939	0.847
LEP	1159	0.817	1161	0.832	1161	0.810
LEP in first year of school	14	0.812	55	0.820	13	0.834
LEP exited in past 1 year	203	0.767	202	0.838	203	0.773
LEP exited in past 2 years	116	0.812	116	0.850	115	0.768
<b>TITLE I</b>						
No	15441	0.875	15468	0.894	15445	0.851
Yes	3980	0.879	4009	0.890	3986	0.854
<b>MIGRANT STATUS</b>						
No	19129	0.877	19180	0.895	19139	0.853
Yes	292	0.835	297	0.848	292	0.837
<b>GIFTED AND TALENTED</b>						
No	18118	0.869	18174	0.887	18128	0.845
Yes	1303	0.731	1303	0.817	1303	0.660
<b>NEGLECTED OR DELINQUENT</b>						
No	19417	0.878	19473	0.896	19427	0.855
Yes	4	N/A	4	N/A	4	N/A
<b>HOMELESS</b>						
No	19392	0.878	19446	0.895	19402	0.855
Yes	29	0.896	31	0.874	29	0.869
<b>HOME SCHOOL</b>						
No	19421	0.878	19477	0.896	19431	0.855
Yes	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>						
No	17378	0.861	17427	0.884	17382	0.831
Yes	1803	0.843	1809	0.852	1808	0.811
SPE exited in past 1 year	166	0.870	167	0.873	167	0.839
SPE exited in past 2 years	74	0.884	74	0.883	74	0.868
<b>504 PLAN</b>						
No	19182	0.878	19237	0.896	19191	0.855
Yes	239	0.877	240	0.883	240	0.842

**Table 12.10. Grade 7: Reliability**

	Reading		Mathematics		Lang. Usage		Science	
	N	Alpha	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19683	0.897	19714	0.912	19682	0.854	19587	0.880
<b>GENDER</b>								
Male	10182	0.901	10208	0.917	10186	0.856	10136	0.892
Female	9501	0.892	9506	0.907	9496	0.843	9451	0.864
<b>ETHNICITY</b>								
American Indian/Alaskan Native	322	0.875	322	0.888	321	0.814	312	0.843
Asian	251	0.901	255	0.918	253	0.853	252	0.884
Black/African American	210	0.889	207	0.911	209	0.862	207	0.846
Hawaiian/Other Pacific Islander	58	0.840	58	0.871	58	0.745	58	0.844
White	15901	0.892	15911	0.910	15901	0.846	15845	0.875
Hispanic or Latino	2612	0.874	2632	0.885	2612	0.827	2588	0.845
Other/Unknown	329	0.893	329	0.899	328	0.845	325	0.888
<b>FREE OR REDUCED LUNCH</b>								
No	11858	0.885	11858	0.906	11855	0.838	11805	0.871
Yes	7825	0.890	7856	0.902	7827	0.844	7782	0.869
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18245	0.893	18247	0.910	18241	0.846	18160	0.876
LEP	1140	0.832	1144	0.851	1143	0.796	1131	0.779
LEP in first year of school	32	0.788	58	0.834	33	0.642	32	0.735
LEP exited in past 1 year	139	0.828	138	0.896	139	0.788	138	0.814
LEP exited in past 2 years	127	0.830	127	0.868	126	0.741	126	0.783
<b>TITLE I</b>								
No	16764	0.895	16786	0.912	16769	0.851	16694	0.878
Yes	2919	0.891	2928	0.905	2913	0.844	2893	0.876
<b>MIGRANT STATUS</b>								
No	19414	0.896	19433	0.912	19412	0.852	19318	0.879
Yes	269	0.856	281	0.871	270	0.839	269	0.809
<b>GIFTED AND TALENTED</b>								
No	18602	0.892	18633	0.907	18602	0.845	18511	0.872
Yes	1081	0.726	1081	0.797	1080	0.670	1076	0.762
<b>NEGLECTED OR DELINQUENT</b>								
No	19672	0.897	19702	0.912	19670	0.854	19575	0.880
Yes	11	N/A	12	N/A	12	N/A	12	N/A
<b>HOMELESS</b>								
No	19650	0.897	19680	0.912	19648	0.854	19553	0.880
Yes	33	0.893	34	0.902	34	0.830	34	0.864
<b>HOME SCHOOL</b>								
No	19683	0.897	19714	0.912	19682	0.854	19587	0.880
Yes	0	N/A	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>								
No	17725	0.883	17756	0.903	17723	0.835	17658	0.872
Yes	1749	0.847	1748	0.866	1750	0.787	1723	0.844
SPE exited in past 1 year	143	0.841	144	0.871	143	0.788	140	0.838
SPE exited in past 2 years	66	0.831	66	0.847	66	0.728	66	0.842
<b>504 PLAN</b>								
No	19336	0.897	19367	0.913	19335	0.854	19243	0.880
Yes	347	0.885	347	0.888	347	0.840	344	0.882

**Table 12.11. Grade 8: Reliability**

	Reading		Mathematics		Lang. Usage	
	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19840	0.884	19864	0.901	19821	0.852
<b>GENDER</b>						
Male	10246	0.886	10262	0.906	10238	0.857
Female	9594	0.879	9602	0.895	9583	0.840
<b>ETHNICITY</b>						
American Indian/Alaskan Native	310	0.877	310	0.864	309	0.833
Asian	238	0.892	239	0.897	238	0.850
Black/African American	179	0.887	179	0.898	179	0.875
Hawaiian/Other Pacific Islander	68	0.851	68	0.886	68	0.821
White	16093	0.876	16092	0.898	16077	0.842
Hispanic or Latino	2629	0.875	2651	0.877	2626	0.841
Other/Unknown	323	0.839	325	0.887	324	0.840
<b>FREE OR REDUCED LUNCH</b>						
No	12023	0.869	12032	0.895	12024	0.831
Yes	7817	0.883	7832	0.890	7797	0.851
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	18538	0.878	18525	0.898	18525	0.844
LEP	996	0.832	999	0.835	993	0.811
LEP in first year of school	33	0.894	69	0.878	32	0.860
LEP exited in past 1 year	145	0.827	144	0.870	144	0.793
LEP exited in past 2 years	128	0.832	127	0.845	127	0.810
<b>TITLE I</b>						
No	17405	0.881	17418	0.900	17391	0.848
Yes	2435	0.883	2446	0.886	2430	0.850
<b>MIGRANT STATUS</b>						
No	19582	0.883	19597	0.900	19564	0.851
Yes	258	0.854	267	0.865	257	0.838
<b>GIFTED AND TALENTED</b>						
No	18716	0.878	18740	0.894	18697	0.845
Yes	1124	0.735	1124	0.820	1124	0.698
<b>NEGLECTED OR DELINQUENT</b>						
No	19815	0.884	19838	0.900	19795	0.852
Yes	25	0.890	26	0.861	26	0.893
<b>HOMELESS</b>						
No	19812	0.884	19835	0.901	19791	0.852
Yes	28	0.886	29	0.861	30	0.883
<b>HOME SCHOOL</b>						
No	19840	0.884	19864	0.901	19821	0.852
Yes	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>						
No	17952	0.868	17980	0.891	17937	0.829
Yes	1717	0.832	1712	0.835	1712	0.805
SPE exited in past 1 year	105	0.865	107	0.888	106	0.837
SPE exited in past 2 years	66	0.865	65	0.881	66	0.839
<b>504 PLAN</b>						
No	19484	0.884	19508	0.901	19464	0.852
Yes	356	0.881	356	0.886	357	0.847

**Table 12.12. Grade 9: Reliability**

	Reading		Mathematics		Lang. Usage	
	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	20544	0.839	20611	0.904	20523	0.814
<b>GENDER</b>						
Male	10645	0.849	10676	0.910	10637	0.817
Female	9899	0.823	9935	0.898	9886	0.803
<b>ETHNICITY</b>						
American Indian/Alaskan Native	322	0.832	321	0.893	321	0.766
Asian	243	0.826	252	0.911	242	0.803
Black/African American	193	0.831	201	0.904	194	0.786
Hawaiian/Other Pacific Islander	65	0.785	65	0.895	64	0.789
White	16773	0.825	16789	0.901	16763	0.806
Hispanic or Latino	2585	0.832	2619	0.874	2579	0.756
Other/Unknown	363	0.819	364	0.897	360	0.816
<b>FREE OR REDUCED LUNCH</b>						
No	12877	0.815	12904	0.899	12870	0.804
Yes	7667	0.842	7707	0.896	7653	0.796
<b>LIMITED ENGLISH PROFICIENCY</b>						
Not LEP	19361	0.830	19366	0.902	19343	0.808
LEP	952	0.802	958	0.829	950	0.684
LEP in first year of school	31	0.858	87	0.890	30	0.830
LEP exited in past 1 year	113	0.737	113	0.859	113	0.721
LEP exited in past 2 years	87	0.780	87	0.865	87	0.702
<b>TITLE I</b>						
No	19119	0.836	19185	0.904	19108	0.813
Yes	1425	0.847	1426	0.886	1415	0.803
<b>MIGRANT STATUS</b>						
No	20281	0.836	20341	0.904	20261	0.812
Yes	263	0.811	270	0.867	262	0.714
<b>GIFTED AND TALENTED</b>						
No	19603	0.835	19668	0.899	19581	0.801
Yes	941	0.563	943	0.805	942	0.713
<b>NEGLECTED OR DELINQUENT</b>						
No	20477	0.839	20542	0.904	20454	0.813
Yes	67	0.774	69	0.844	69	0.719
<b>HOMELESS</b>						
No	20522	0.839	20589	0.904	20501	0.814
Yes	22	0.703	22	0.924	22	0.678
<b>HOME SCHOOL</b>						
No	20544	0.839	20611	0.904	20523	0.814
Yes	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>						
No	18669	0.814	18739	0.895	18649	0.798
Yes	1735	0.805	1732	0.848	1735	0.703
SPE exited in past 1 year	92	0.768	92	0.844	92	0.715
SPE exited in past 2 years	48	0.821	48	0.875	47	0.712
<b>504 PLAN</b>						
No	20176	0.838	20244	0.905	20156	0.814
Yes	368	0.854	367	0.892	367	0.787

**Table 12.13. Grade 10: Reliability**

	Reading		Mathematics		Lang. Usage		Science	
	N	Alpha	N	Alpha	N	Alpha	N	Alpha
<b>ALL STUDENTS</b>	19092	0.895	19131	0.891	19101	0.827	18769	0.899
<b>GENDER</b>								
Male	9799	0.900	9816	0.901	9796	0.832	9618	0.911
Female	9293	0.888	9315	0.880	9305	0.813	9151	0.884
<b>ETHNICITY</b>								
American Indian/Alaskan Native	249	0.893	246	0.879	249	0.818	225	0.891
Asian	219	0.902	227	0.905	219	0.825	216	0.909
Black/African American	178	0.886	178	0.901	175	0.830	168	0.892
Hawaiian/Other Pacific Islander	80	0.884	80	0.871	80	0.774	77	0.892
White	15883	0.889	15881	0.888	15891	0.819	15640	0.896
Hispanic or Latino	2141	0.880	2175	0.862	2144	0.806	2112	0.865
Other/Unknown	342	0.889	344	0.887	343	0.800	331	0.883
<b>FREE OR REDUCED LUNCH</b>								
No	12959	0.886	12970	0.886	12965	0.815	12737	0.895
Yes	6133	0.892	6161	0.884	6136	0.822	6032	0.892
<b>LIMITED ENGLISH PROFICIENCY</b>								
Not LEP	18095	0.890	18100	0.889	18104	0.820	17791	0.897
LEP	810	0.834	808	0.841	811	0.759	792	0.793
LEP in first year of school	17	0.722	52	0.850	15	0.724	16	0.639
LEP exited in past 1 year	103	0.870	103	0.828	103	0.728	102	0.903
LEP exited in past 2 years	67	0.869	68	0.822	68	0.721	68	0.883
<b>TITLE I</b>								
No	18380	0.894	18414	0.891	18389	0.825	18064	0.899
Yes	712	0.893	717	0.863	712	0.820	705	0.876
<b>MIGRANT STATUS</b>								
No	18881	0.894	18914	0.891	18889	0.826	18559	0.899
Yes	211	0.847	217	0.842	212	0.790	210	0.808
<b>GIFTED AND TALENTED</b>								
No	18312	0.891	18352	0.888	18322	0.821	17997	0.895
Yes	780	0.741	779	0.780	779	0.687	772	0.823
<b>NEGLECTED OR DELINQUENT</b>								
No	19008	0.895	19047	0.891	19018	0.827	18689	0.899
Yes	84	0.853	84	0.866	83	0.767	80	0.840
<b>HOMELESS</b>								
No	19074	0.895	19113	0.891	19083	0.827	18752	0.899
Yes	18	0.930	18	0.898	18	0.857	17	0.908
<b>HOME SCHOOL</b>								
No	19092	0.895	19131	0.891	19101	0.827	18769	0.899
Yes	0	N/A	0	N/A	0	N/A	0	N/A
<b>SPECIAL EDUCATION</b>								
No	17443	0.882	17479	0.878	17449	0.809	17198	0.894
Yes	1550	0.846	1552	0.860	1551	0.750	1472	0.840
SPE exited in past 1 year	67	0.857	67	0.879	68	0.772	66	0.836
SPE exited in past 2 years	32	0.849	33	0.877	33	0.789	33	0.869
<b>504 PLAN</b>								
No	18765	0.895	18803	0.892	18775	0.828	18452	0.900
Yes	327	0.888	328	0.866	326	0.786	317	0.885

## 12.1 Standard Error of Measurement

The standard error of measurement (SEM) uses the information from the test along with an estimate of reliability to make statements about the degree to which error affects individual scores. It is based on the premise that underlying traits, such as academic achievement, cannot be measured exactly. The standard error expresses unreliability in the raw score metric. Using the standard error of measurement, an error band can be placed around an individual score indicating the degree to which error might be affecting that score. In true-score test theory, the SEM can be calculated by:

$$SEM = \sigma_X \sqrt{1 - \rho_{XX}}$$

where  $\sigma_X$  is the standard deviation of the total test (observed measure scores), and  $\rho_{XX}$  is the Coefficient Alpha reliability estimate for the test.

Table 12.14 shows the traditional SEMs after converting them from the raw score metric into the scale score metric.

**Table 12.14. Standard Error of Measurement (SEM) by Grade and Content**

Grade	Reading	Math	Language Usage	Science
3	3.892	4.397	4.658	
4	3.925	4.122	4.197	
5	3.754	3.992	3.957	3.765
6	3.815	3.901	4.004	
7	3.989	3.908	3.648	3.437
8	3.691	3.831	3.809	
9	3.734	3.803	3.608	
10	3.818	3.911	3.636	3.426

The true-score test theory approach to judging a test's consistency (SEM) can be useful for making overall comparisons between alternate forms. However, it is not very useful for judging the precision with which a specific student's score is known. A more useful measure within the Rasch measurement model is the asymptotic or conditional standard error that pertain to each unique ability estimate. In general, ability estimates from scores near the center of the test are known with greater precision than are abilities associated with extremely high or low scores.

Specific to criterion-referenced tests, like the ISAT, these conditional standards errors (CSEs) are most informative at the decision points (aka cutpoints) that define whether students are placed into Below Basic, Basic, Proficient, or Advanced. These values are shown in Tables 12.15–12.18 in the scale score metric. In most cases, the CSEs at the Basic/Below Basic cutpoint and the Proficient/Basic cutpoint are lower than the traditional SEM values. However, the Advanced/Proficient values are somewhat higher. This is due to less precise measurement at the extremes of the score distribution. The complete set of conditional standard errors for every obtainable score can be found in Appendix U as part of the raw to scale score conversions.

**Table 12.15. Reading Conditional Standard Error of Measurement at the Cut Points**

<b>Content</b>	<b>Grade</b>	<b>Cut Point</b>	<b>CSEM</b>
Reading	3	Basic/Below Basic	3.52
Reading	3	Proficient/Basic	3.42
Reading	3	Advanced/Proficient	4.18
Reading	4	Basic/Below Basic	3.54
Reading	4	Proficient/Basic	3.40
Reading	4	Advanced/Proficient	3.95
Reading	5	Basic/Below Basic	3.39
Reading	5	Proficient/Basic	3.27
Reading	5	Advanced/Proficient	4.08
Reading	6	Basic/Below Basic	3.38
Reading	6	Proficient/Basic	3.30
Reading	6	Advanced/Proficient	4.10
Reading	7	Basic/Below Basic	3.37
Reading	7	Proficient/Basic	3.30
Reading	7	Advanced/Proficient	4.13
Reading	8	Basic/Below Basic	3.41
Reading	8	Proficient/Basic	3.15
Reading	8	Advanced/Proficient	3.72
Reading	9	Basic/Below Basic	3.52
Reading	9	Proficient/Basic	3.26
Reading	9	Advanced/Proficient	3.71
Reading	10	Basic/Below Basic	3.29
Reading	10	Proficient/Basic	3.15
Reading	10	Advanced/Proficient	4.01



**Table 12.16. Mathematics Conditional Standard Error of Measurement at the Cut Points**

<b>Content</b>	<b>Grade</b>	<b>Cut Point</b>	<b>SEM</b>
Mathematics	3	Basic/Below Basic	3.53
Mathematics	3	Proficient/Basic	3.39
Mathematics	3	Advanced/Proficient	4.35
Mathematics	4	Basic/Below Basic	3.50
Mathematics	4	Proficient/Basic	3.44
Mathematics	4	Advanced/Proficient	4.45
Mathematics	5	Basic/Below Basic	3.44
Mathematics	5	Proficient/Basic	3.34
Mathematics	5	Advanced/Proficient	3.97
Mathematics	6	Basic/Below Basic	3.33
Mathematics	6	Proficient/Basic	3.18
Mathematics	6	Advanced/Proficient	3.88
Mathematics	7	Basic/Below Basic	3.28
Mathematics	7	Proficient/Basic	3.20
Mathematics	7	Advanced/Proficient	4.07
Mathematics	8	Basic/Below Basic	3.37
Mathematics	8	Proficient/Basic	3.24
Mathematics	8	Advanced/Proficient	3.93
Mathematics	9	Basic/Below Basic	3.57
Mathematics	9	Proficient/Basic	3.35
Mathematics	9	Advanced/Proficient	3.60
Mathematics	10	Basic/Below Basic	3.24
Mathematics	10	Proficient/Basic	3.24
Mathematics	10	Advanced/Proficient	4.06

**Table 12.17. Language Usage Conditional Standard Error of Measurement at the Cut Points**

<b>Content</b>	<b>Grade</b>	<b>Cut Point</b>	<b>SEM</b>
Language Usage	3	Basic/Below Basic	3.75
Language Usage	3	Proficient/Basic	4.27
Language Usage	3	Advanced/Proficient	5.60
Language Usage	4	Basic/Below Basic	3.85
Language Usage	4	Proficient/Basic	3.80
Language Usage	4	Advanced/Proficient	4.22
Language Usage	5	Basic/Below Basic	3.42
Language Usage	5	Proficient/Basic	3.55
Language Usage	5	Advanced/Proficient	4.52
Language Usage	6	Basic/Below Basic	3.43
Language Usage	6	Proficient/Basic	3.53
Language Usage	6	Advanced/Proficient	4.66
Language Usage	7	Basic/Below Basic	3.25
Language Usage	7	Proficient/Basic	3.31
Language Usage	7	Advanced/Proficient	4.56
Language Usage	8	Basic/Below Basic	3.26
Language Usage	8	Proficient/Basic	3.46
Language Usage	8	Advanced/Proficient	4.91
Language Usage	9	Basic/Below Basic	3.33
Language Usage	9	Proficient/Basic	3.31
Language Usage	9	Advanced/Proficient	4.39
Language Usage	10	Basic/Below Basic	3.34
Language Usage	10	Proficient/Basic	3.34
Language Usage	10	Advanced/Proficient	4.58

**Table 12.18. Science Conditional Standard Error of Measurement at the Cut Points**

Content	Grade	Cut Point	SEM
Science	5	Basic/Below Basic	3.47
Science	5	Proficient/Basic	3.41
Science	5	Advanced/Proficient	4.11
Science	7	Basic/Below Basic	3.05
Science	7	Proficient/Basic	3.17
Science	7	Advanced/Proficient	3.57
Science	10	Basic/Below Basic	3.02
Science	10	Proficient/Basic	3.08
Science	10	Advanced/Proficient	3.85

## 12.2 Indicators of Consistency

When criterion-referenced tests are used to place the students into two or more performance classifications, it is useful to have some indication of how consistent such classifications are. Decision consistency indices can provide information about accuracy of placement over and above the information provided by conditional standard errors.

### Decision Consistency Index

To solve the problem of a complex assessment, Livingston and Lewis (1995) proposed an effective test length,

$$n = \frac{(\mu_x - X_{\min})(X_{\max} - \mu_x) - r\sigma_x^2}{\sigma_x^2(1 - r)},$$

which transforms the original raw score random variable from  $X = 0, \dots, K$  into a new random variable  $X' = 0, \dots, n$ , where  $n$  is the number of dichotomous, locally independent, equally difficult items required to produce a raw score of the same reliability. Then, using the transformed observed distribution  $X'$ , parameters are estimated for a four parameter beta-binomial model where the conditional error distribution is assumed to be binomial. The  $X'$  distribution is then converted back onto the original  $X$  scale using interpolation. It is a single administration decision consistency method that estimates the consistency of the decisions on alternate forms of a test and the accuracy of those decisions relative to the examinees' true score. It is designed only to estimate a contingency table, not a full bivariate distribution, which means the probability of a consistent decision by chance, and subsequently kappa, cannot be estimated.

The results of the consistency analyses are presented in Table 12.19.

**Table 12.19. Decision Consistency Indices**

<b>Reliability of Performance Levels for the 2007 ISAT</b>			
		<b>Decision Consistency Index</b>	
<b>Grade</b>	<b>Content Area</b>	<b>Two Performance Levels (Proficient versus Not Proficient)</b>	<b>Four Performance Levels</b>
<b>3</b>	<b>Reading</b>	0.91	0.72
	<b>Mathematics</b>	0.93	0.76
	<b>Language Usage</b>	0.86	0.59
<b>4</b>	<b>Reading</b>	0.89	0.71
	<b>Mathematics</b>	0.91	0.70
	<b>Language Usage</b>	0.86	0.65
<b>5</b>	<b>Reading</b>	0.89	0.69
	<b>Mathematics</b>	0.88	0.70
	<b>Language Usage</b>	0.85	0.64
	<b>Science</b>	0.81	0.65
<b>6</b>	<b>Reading</b>	0.89	0.70
	<b>Mathematics</b>	0.88	0.71
	<b>Language Usage</b>	0.85	0.64
<b>7</b>	<b>Reading</b>	0.89	0.71
	<b>Mathematics</b>	0.89	0.72
	<b>Language Usage</b>	0.84	0.67
	<b>Science</b>	0.86	0.64
<b>8</b>	<b>Reading</b>	0.92	0.74
	<b>Mathematics</b>	0.89	0.71
	<b>Language Usage</b>	0.84	0.67
<b>9</b>	<b>Reading</b>	0.92	0.68
	<b>Mathematics</b>	0.90	0.73
	<b>Language Usage</b>	0.81	0.66
<b>10</b>	<b>Reading</b>	0.90	0.72
	<b>Mathematics</b>	0.88	0.69
	<b>Language Usage</b>	0.83	0.69
	<b>Science</b>	0.87	0.68

## 13. Performance Levels

Tables 13.1 to 13.4 show the percent of students who were categorized as Below Basic, Basic, Proficient, and Advanced based on their 2007 Spring ISAT scores. The percentages that combine the Proficient and Advanced levels are of particular note as they are used in calculating Adequate Yearly Progress as part of NCLB. The tables are presented by grade for each content area. The process used to establish the cutpoints that define how students are categorized may be found in Appendix W.

**Table 13.1. Reading Performance Level Percentages**

Content	Grade	Percentage of Students			Below Basic
		Advanced	Proficient	Basic	
Reading	3	37.3	43.5	10.1	9.1
Reading	4	30.0	50.6	10.5	8.9
Reading	5	31.5	47.0	11.9	9.6
Reading	6	27.0	50.4	14.1	8.5
Reading	7	32.0	45.0	15.3	7.6
Reading	8	37.9	47.9	10.7	3.5
Reading	9	32.9	54.7	9.4	3.0
Reading	10	31.0	47.8	14.7	6.6

**Table 13.2. Mathematics Performance Level Percentages**

Content	Grade	Percentage of Students			Below Basic
		Advanced	Proficient	Basic	
Mathematics	3	49.2	37.1	11.2	2.4
Mathematics	4	33.3	48.7	13.3	4.8
Mathematics	5	32.0	41.0	19.8	7.2
Mathematics	6	31.0	43.7	19.1	6.1
Mathematics	7	28.9	41.1	19.3	10.6
Mathematics	8	29.2	42.5	20.4	7.9
Mathematics	9	35.8	41.6	14.2	8.4
Mathematics	10	28.5	44.2	17.2	10.2

**Table 13.3. Language Usage Performance Level Percentages**

Content	Grade	Percentage of Students			Below Basic
		Advanced	Proficient	Basic	
Language Usage	3	31.7	34.8	20.6	12.8
Language Usage	4	29.9	50.1	14.6	5.4
Language Usage	5	18.5	50.3	22.1	9.1
Language Usage	6	19.0	48.4	21.9	10.6
Language Usage	7	11.6	53.3	24.8	10.3
Language Usage	8	12.1	49.7	26.0	12.2
Language Usage	9	5.3	53.6	29.0	12.1
Language Usage	10	5.5	58.7	24.2	11.7

**Table 13.4. Science Performance Level Percentages**

<b>Content</b>	<b>Grade</b>	<b>Percentage of Students</b>			
		<b>Advanced</b>	<b>Proficient</b>	<b>Basic</b>	<b>Below Basic</b>
Science	5	13.9	39.5	40.5	6.1
Science	7	24.6	23.1	25.0	27.4
Science	10	18.4	39.7	17.3	24.6

## 14. Linking

Consistent with the RFP, and in consideration of the discussion at the initial Technical Advisory Committee meeting held on September 12–13, 2006, DRC utilized Idaho-owned NWEA items for linking the Spring 2007 reading, mathematics, and science assessments to the operational scale. However, language usage did not have linking items to the operational scale. These linking items were administered along with the DRC-developed items. The DRC-developed items constituted the core section, administered to all students and used for computing adequate yearly progress (AYP), while the linking items were randomly administered to a subset of the student population, approximately 20 percent, in a section immediately following the core section. In language usage, rather than using linking items, and in place of a computer adaptive test, unique fixed-length forms with no overlapping items were administered at each grade level. Given that all of these items were Idaho-owned and had known item parameters, the raw to scale score tables in the existing vertical Rasch metric were constructed directly utilizing the item parameter estimates established by the former vendor. In determining the final linking item set for language usage, the same steps were followed as those for reading, mathematics, and science (see Appendix S).

### ***Linking to the Operational Scale***

DRC's linking protocol utilized Rasch methodologies that established a scale that was resistant to fluctuations in the student population and minor content and statistical differences in test forms from year to year. The DRC-developed items on the Spring 2007 test forms were linked to the operational scale via approximately 14–15 Idaho state-owned items. These anchor items were chosen jointly by DRC's content and psychometric staff.

The formal link to the operational scale was accomplished via a comparison of DRC's free item calibration and the *a priori* calibration values of the Idaho-owned items. Note that the Idaho-owned estimated item difficulties were presumed to be in the common or baseline metric. This assumption was supported in that, in DRC's analysis of the Spring 2006 item-person data file, the item parameter estimates derived from our calibrations indicated a reasonable match to those values passed to us as part of the materials transition from the former vendor. Details of this process can be found in section 4.1 *Replication Study*. As part of this study, DRC excluded from consideration any item where there was a statistically significant difference in the calibrated values.

Table 14.1 shows an example of the linking for mathematics grade 5. It contains an item identifier, the content code (match to Idaho content standards), the *a priori* item calibration value, the DRC-generated item difficulty, the discrepancy between the *a priori* value and the DRC-calibrated value, the linked value (to the baseline metric), and the robust-z. At the bottom of table, descriptive statistics are presented for the linking set for the *a priori* values and DRC's free-calibrated values:

- Means
- Standard deviations
- Ratio of the standard deviations
- Correlation between the item difficulties
- Additive constant
- Medians

- Inter-quartile range

The steps that were used to accomplish the link can be found in Appendix S, as described in Huynh (2000) as part of the technical documentation for the South Carolina Palmetto Achievement Challenge Tests (PACT). Given that the items were not developed by DRC and were reviewed for content and statistical appropriateness as part of the initial selection of the linking sets, the item exclusions were based on statistical evidence alone. Under more typical circumstances, a review by content specialists would have been done before excluding any item. Following these steps, in the example below, resulted in the removal of two items from the link. The second table shows the same information as the first table with the two items excluded from the computations.

**Table 14.1. Linking Example #1 – Grade 5 Mathematics**

Item ID	Content Code	Rasch Difficulty			Linked to	
		Baseline	ISAT 2007	Discrepancy	Base	Robust Z
514258	1.2.2	1.7	0.119	1.581	1.833	-0.309
514277	1.2.4	-0.4	-1.791	1.391	-0.077	-0.974
514257	1.2.7	3.0	1.329	1.671	3.043	0.005
514259	1.2.7	0.4	-0.556	0.956	1.158	-2.495
514262	3.4.1	1.3	-0.091	1.391	1.623	-0.974
514264	5.2.1	1.9	0.313	1.587	2.027	-0.288
514266	4.3.1	2.1	-0.377	2.477	1.337	2.823
514269	5.1.1	2.9	1.214	1.686	2.928	0.058
514270	2.1.4	3.4	1.132	2.268	2.846	2.093
514273	5.5.1	0.8	-1.058	1.858	0.656	0.659
514275	2.1.4	0.1	-0.546	0.646	1.168	-3.579
514276	4.1.1	0.8	-0.868	1.668	0.846	-0.005
514278	5.3.1	2.5	-0.593	3.093	1.121	4.977
514391	3.4.1	3.5	1.774	1.726	3.488	0.198
<b>Mean</b>		1.714	0.000	1.714	1.714	0.156
<b>SD</b>		1.257	1.037	0.606	1.037	2.119
<b>SD Ratio</b>		1.212				
<b>Correlation</b>		0.878				
<b>Add.</b>						
<b>Constant</b>		1.714				
<b>Median</b>				1.670		
<b>Q</b>				0.387		



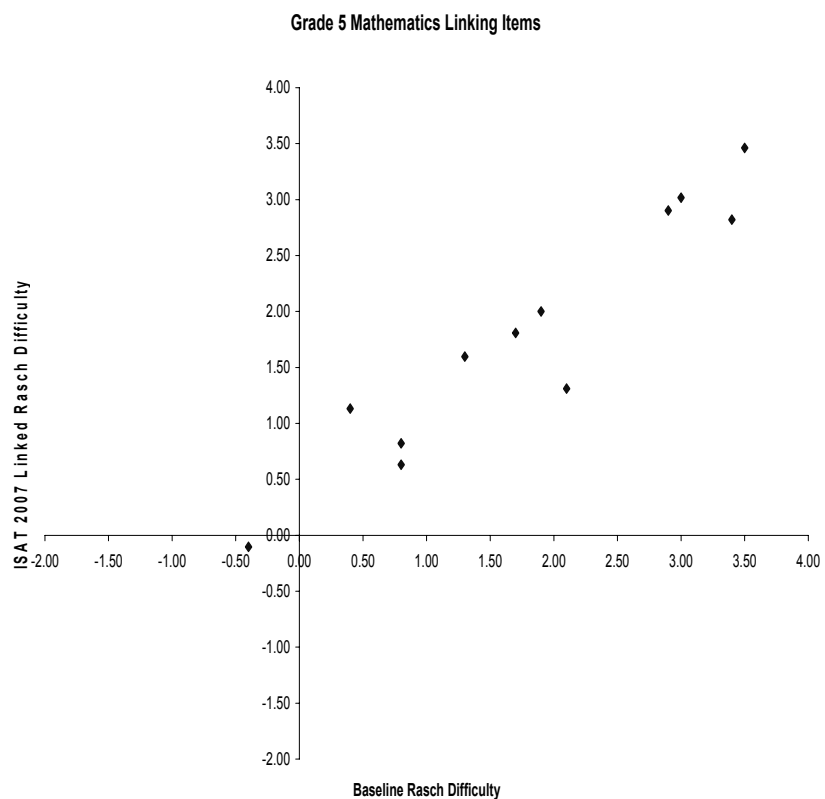
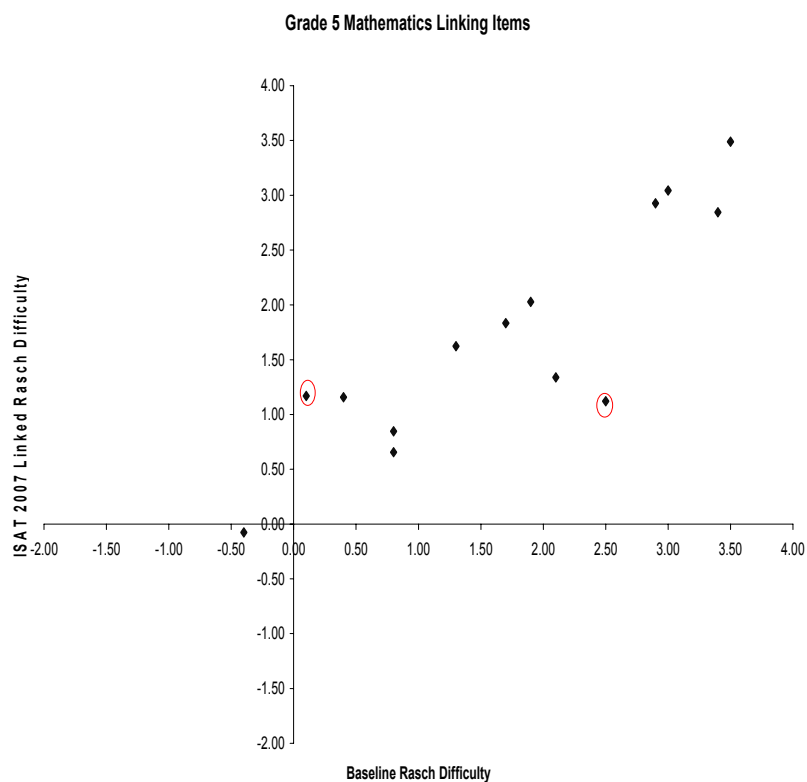
**Table 14.1 (continued). Linking Example #1 – Grade 5 Mathematics**

Two linking items removed.

Item ID	Content Code	Rasch Difficulty			Linked to Base	Robust Z
		Baseline	ISAT 2007	Discrepancy		
514258	1.2.2	1.7	0.025	1.675	1.808	-0.554
514277	1.2.4	-0.4	-1.910	1.510	-0.127	-1.617
514257	1.2.7	3.0	1.247	1.753	3.030	-0.052
514259	1.2.7	0.4	-0.657	1.057	1.126	-4.536
514262	3.4.1	1.3	-0.187	1.487	1.596	-1.765
514264	5.2.1	1.9	0.220	1.680	2.003	-0.522
514266	4.3.1	2.1	-0.476	2.576	1.307	5.251
514269	5.1.1	2.9	1.131	1.769	2.914	0.052
514270	2.1.4	3.4	1.048	2.352	2.831	3.808
514273	5.5.1	0.8	-1.165	1.965	0.618	1.314
514275	2.1.4					
514276	4.1.1	0.8	-0.973	1.773	0.810	0.077
514278	5.3.1					
514391	3.4.1	3.5	1.697	1.803	3.480	0.271
<b>Mean</b>		1.783	0.000	1.783	1.783	0.144
<b>SD</b>		1.253	1.108	0.393	1.108	2.532
<b>SD Ratio</b>		1.130				
<b>Correlation</b>		0.952				
<b>Add.</b>						
<b>Constant</b>		1.783				
<b>Median</b>				1.761		
<b>Q</b>				0.210		

A plot of the two sets of item difficulties is shown in Figure 14.1. Note that the two item exclusions are circled in red. The second graph shows the same plot, but without the two excluded items. The complete set of plots may be found in Appendix T, following the format of the former.

**Figure 14.1**



As noted previously, linking for language usage was different from reading, mathematics, and science. Table 14.2 and Figure 14.2 show an example using grade 7.

**Table 14.2. Linking Example #2 – Grade 7 Language Usage**

45 linking items.

Item ID	Content Code	Rasch Difficulty			Linked to	
		Baseline	ISAT 2007	Discrepancy	Base	Robust Z
510608	5.3.2	-0.7	-1.141	0.441	-0.041	-1.654
510723	5.2.3	0.8	-0.383	1.183	0.717	-0.037
509881	5.3.2	0.7	0.431	0.269	1.531	-2.029
510160	5.3.2	-2.0	-0.957	-1.043	0.143	-4.889
510754	5.4.1	1.2	0.000	1.200	1.100	0.000
510604	5.4.2	0.3	-0.546	0.846	0.554	-0.772
510777	5.4.2	1.5	-0.141	1.641	0.959	0.961
510534	3.3.4	0.2	-1.293	1.493	-0.193	0.639
510589	3.1.3	1.2	-0.220	1.420	0.880	0.480
510547	3.3.2	-0.7	-2.209	1.509	-1.109	0.673
509937	3.3.2	-0.1	-0.712	0.612	0.388	-1.282
510520	3.3.4	1.6	0.405	1.195	1.505	-0.011
510499	3.3.1	1.3	-0.230	1.530	0.870	0.719
509723	3.4.1	-0.3	-1.714	1.414	-0.614	0.466
510758	5.4.1	3.0	1.810	1.190	2.910	-0.022
510691	3.3.1	0.7	-0.663	1.363	0.437	0.355
510738	5.4.2	1.5	0.934	0.566	2.034	-1.382
509913	5.3.2	0.5	-0.936	1.436	0.164	0.514
510591	5.3.1	2.0	0.453	1.547	1.553	0.756
510352	3.3.3	2.4	0.926	1.474	2.026	0.597
510275	5.3.1	2.7	0.993	1.707	2.093	1.105
510677	3.1.2	1.4	0.038	1.362	1.138	0.353
509927	3.1.2	2.3	0.952	1.348	2.052	0.323
510752	3.3.2	3.1	1.318	1.782	2.418	1.269
509922	3.4.1	0.4	-0.936	1.336	0.164	0.296
510535	5.2.3	0.7	0.059	0.641	1.159	-1.218
510799	3.4.1	0.3	-0.403	0.703	0.697	-1.083
510795	5.3.1	2.7	1.502	1.198	2.602	-0.004
510709	3.3.1	0.3	-0.906	1.206	0.194	0.013
510618	5.3.2	2.7	0.496	2.204	1.596	2.188
510737	5.4.2	2.2	0.916	1.284	2.016	0.183
510672	3.1.4	0.0	-1.012	1.012	0.088	-0.410
510532	3.1.4	2.1	1.045	1.055	2.145	-0.316
510636	3.1.4	0.9	-0.049	0.949	1.051	-0.547
510657	3.1.3	1.3	-0.045	1.345	1.055	0.316
509934	3.1.2	2.9	1.153	1.747	2.253	1.192
510560	3.1.2	0.3	-0.489	0.789	0.611	-0.896
510705	3.1.3	1.0	0.303	0.697	1.403	-1.096
510503	3.3.3	1.8	0.554	1.246	1.654	0.100
510734	3.3.4	2.0	0.738	1.262	1.838	0.135
510537	3.3.1	1.6	0.457	1.143	1.557	-0.124
510722	5.2.1	0.9	-0.063	0.963	1.037	-0.517
510742	5.4.2	0.8	0.110	0.690	1.210	-1.112

Table 14.2 (continued). Linking Example #2 – Grade 7 Language Usage

Item ID	Content Code	Rasch Difficulty			Linked to	
		Baseline	ISAT 2007	Discrepancy	Base	Robust Z
510641	5.3.2	0.2	0.454	-0.254	1.554	-3.169
510689	3.3.1	-0.2	-1.000	0.800	0.100	-0.872
<b>Mean</b>		1.100	0.000	1.100	1.100	-0.218
<b>SD</b>		1.120	0.886	0.553	0.886	1.205
<b>SD Ratio</b>		1.263				
<b>Correlation</b>		0.873				
<b>Add.</b>						
<b>Constant</b>		1.100				
<b>Median</b>				1.200		
<b>Q</b>				0.620		

Two linking items removed.

Item ID	Content Code	Rasch Difficulty			Linked to	
		Baseline	ISAT 2007	Discrepancy	Base	Robust Z
510608	5.3.2	-0.7	-1.155	0.455	0.038	-1.946
510723	5.2.3	0.8	-0.395	1.195	0.798	-0.061
509881	5.3.2	0.7	0.421	0.279	1.614	-2.394
510160	5.3.2					
510754	5.4.1	1.2	-0.012	1.212	1.181	-0.018
510604	5.4.2	0.3	-0.559	0.859	0.634	-0.917
510777	5.4.2	1.5	-0.153	1.653	1.040	1.106
510534	3.3.4	0.2	-1.308	1.508	-0.115	0.736
510589	3.1.3	1.2	-0.231	1.431	0.962	0.540
510547	3.3.2	-0.7	-2.225	1.525	-1.032	0.779
509937	3.3.2	-0.1	-0.725	0.625	0.468	-1.513
510520	3.3.4	1.6	0.394	1.206	1.587	-0.033
510499	3.3.1	1.3	-0.242	1.542	0.951	0.823
509723	3.4.1	-0.3	-1.730	1.430	-0.537	0.537
510758	5.4.1	3.0	1.799	1.201	2.992	-0.046
510691	3.3.1	0.7	-0.675	1.375	0.518	0.397
510738	5.4.2	1.5	0.924	0.576	2.117	-1.638
509913	5.3.2	0.5	-0.949	1.449	0.244	0.586
510591	5.3.1	2.0	0.442	1.558	1.635	0.864
510352	3.3.3	2.4	0.916	1.484	2.109	0.675
510275	5.3.1	2.7	0.983	1.717	2.176	1.269
510677	3.1.2	1.4	0.027	1.373	1.220	0.392
509927	3.1.2	2.3	0.942	1.358	2.135	0.354
510752	3.3.2	3.1	1.308	1.792	2.501	1.460
509922	3.4.1	0.4	-0.949	1.349	0.244	0.331
510535	5.2.3	0.7	0.047	0.653	1.240	-1.442
510799	3.4.1	0.3	-0.415	0.715	0.778	-1.284
510795	5.3.1	2.7	1.492	1.208	2.685	-0.028
510709	3.3.1	0.3	-0.919	1.219	0.274	0.000

Table 14.2 (continued). Linking Example #2 – Grade 7 Language Usage

Item ID	Content Code	Rasch Difficulty			Linked to	
		Baseline	ISAT 2007	Discrepancy	Base	Robust Z
510618	5.3.2	2.7	0.485	2.215	1.678	2.537
510737	5.4.2	2.2	0.906	1.294	2.099	0.191
510672	3.1.4	0.0	-1.026	1.026	0.167	-0.492
510532	3.1.4	2.1	1.035	1.065	2.228	-0.392
510636	3.1.4	0.9	-0.060	0.960	1.133	-0.660
510657	3.1.3	1.3	-0.057	1.357	1.136	0.352
509934	3.1.2	2.9	1.143	1.757	2.336	1.370
510560	3.1.2	0.3	-0.501	0.801	0.692	-1.065
510705	3.1.3	1.0	0.292	0.708	1.485	-1.302
510503	3.3.3	1.8	0.544	1.256	1.737	0.094
510734	3.3.4	2.0	0.728	1.272	1.921	0.135
510537	3.3.1	1.6	0.446	1.154	1.639	-0.166
510722	5.2.1	0.9	-0.074	0.974	1.119	-0.624
510742	5.4.2	0.8	0.099	0.701	1.292	-1.320
510641	5.3.2					
510689	3.3.1	-0.2	-1.013	0.813	0.180	-1.034
<b>Mean</b>		1.193	0.000	1.193	1.193	-0.066
<b>SD</b>		1.028	0.894	0.401	0.894	1.020
<b>SD Ratio</b>		1.150				
<b>Correlation</b>		0.922				
<b>Add.</b>						
<b>Constant</b>		1.193				
<b>Median</b>				1.219		
<b>Q</b>				0.531		

**Figure 14.2**

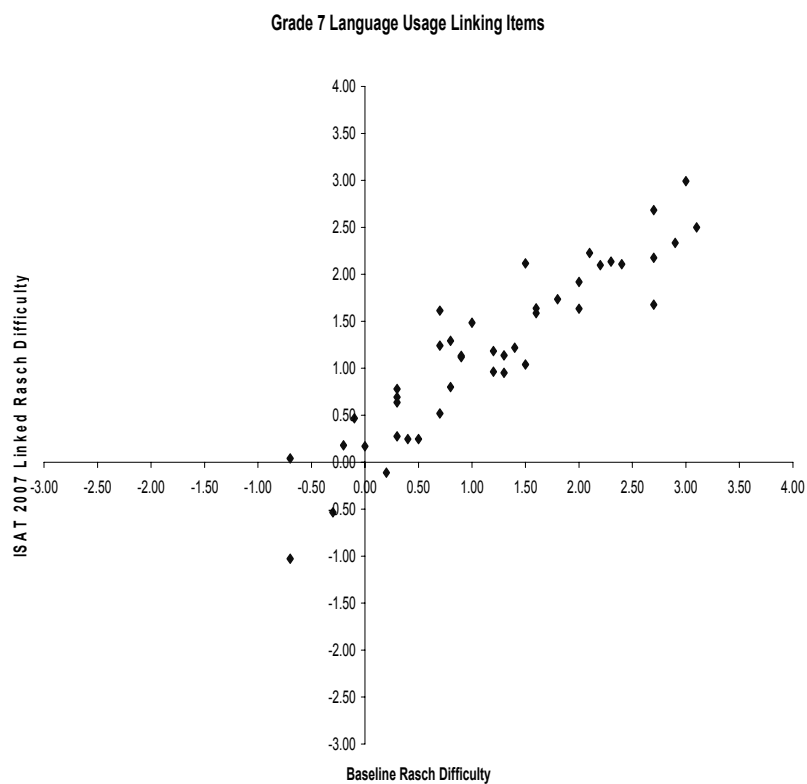
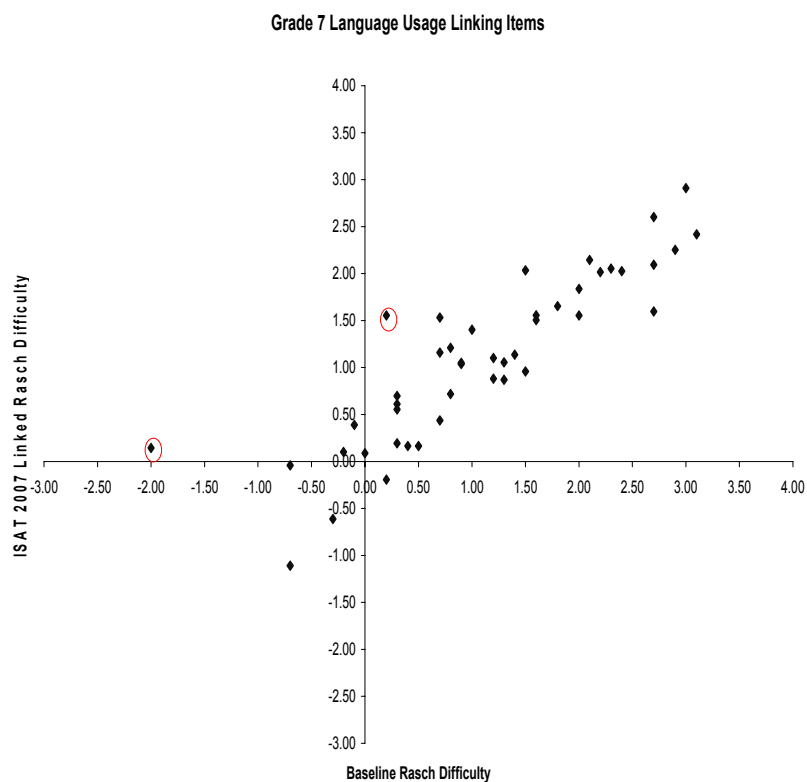


Table 14.3 shows the summary information for the complete set of links. Relevant information therein is the number of items that were eligible for deletion, the number deleted, the resulting ratio of the item difficulties, the correlation between the item difficulties, and the additive constant that was applied to the DRC-calibrated values to link to the operational scale.



**Table 14.3. Linking Summary**

Content	Grade	Number of Links			SD Ratio	Correlation	Additive Constant
		Total	Eligible for Deletion	Deleted			
Reading	3	14	0	0	1.324	0.934	-0.007
Reading	4	14	2	1	1.049	0.973	0.423
Reading	5	14	1	1	0.981	0.748	0.792
Reading	6	14	0	0	1.850	0.652	1.243
Reading	7	14	2	0	1.108	0.900	1.814
Reading	8	14	1	1	0.998	0.940	2.108
Reading	10	45	6	2	1.323	0.911	1.998
Mathematics	3	14	4	1	1.312	0.819	2.823
Mathematics	4	14	3	2	0.947	0.985	-0.167
Mathematics	5	14	0	0	0.970	0.887	0.850
Mathematics	6	14	5	2	1.130	0.952	1.783
Mathematics	7	14	2	1	1.218	0.848	2.277
Mathematics	8	14	2	2	1.135	0.955	3.058
Mathematics	10	14	1	0	1.087	0.883	3.564
Lang. Use.	3	45	6	3	1.205	0.922	3.631
Lang. Use.	4	14	2	2	1.112	0.955	4.558
Lang. Use.	5	40	10	4	0.971	0.964	-1.881
Lang. Use.	6	40	11	5	1.087	0.972	0.069
Lang. Use.	7	42	7	2	1.200	0.957	0.338
Lang. Use.	8	42	7	2	1.086	0.960	0.668
Lang. Use.	10	45	5	2	1.150	0.922	1.193
Science	5	45	11	4	1.060	0.940	1.254
Science	7	45	9	4	1.264	0.920	1.990
Science	10	45	4	2	1.094	0.918	2.086

## 15. Scaling

To allow for ease of interpretation, the logit scale is transformed into scale score units. The equation used to translate the logit abilities to the scale score metric is 10 times the logit plus 200. This eliminates the confusion of having negative scores that may carry equally negative connotation. The 10 represents the standard deviation of the scale and the 200 represents the origin or center. Given that the ISAT is a vertically developed scale, this equation is applied for all grades. As a convention, the same equation is used to translate reading, mathematics, language usage, and science.

### ***15.1 Raw to Scale Score Conversions***

Table 15.1 is an example of the raw score to Rasch logit ability to scale score conversion for Grade 4 Mathematics. In addition, the conditional standard error and actual student frequencies for each raw score are shown. The complete set of conversion tables can be found in Appendix U.

**Table 15.1. Grade 4 Mathematics Core Test: Raw to Scale Conversion**

Raw Score	Ability	Scale Score	SEM	Frequency
0	-5.329	147	18.37	0
1	-4.096	159	10.21	0
2	-3.360	166	7.36	0
3	-2.912	171	6.13	0
4	-2.581	174	5.42	2
5	-2.314	177	4.94	6
6	-2.087	179	4.60	6
7	-1.888	181	4.34	20
8	-1.708	183	4.14	39
9	-1.543	185	3.99	69
10	-1.389	186	3.86	92
11	-1.244	188	3.75	128
12	-1.107	189	3.67	164
13	-0.975	190	3.60	193
14	-0.847	192	3.54	222
15	-0.723	193	3.50	257
16	-0.602	194	3.47	316
17	-0.483	195	3.44	328
18	-0.365	196	3.42	349
19	-0.248	198	3.42	430
20	-0.131	199	3.42	465
21	-0.014	200	3.42	471
22	0.104	201	3.44	518
23	0.223	202	3.46	603
24	0.344	203	3.49	660
25	0.467	205	3.54	715
26	0.594	206	3.59	791
27	0.725	207	3.65	843
28	0.861	209	3.73	923
29	1.003	210	3.82	1036
30	1.153	212	3.93	1104
31	1.313	213	4.07	1177
32	1.486	215	4.24	1208
33	1.674	217	4.45	1271
34	1.883	219	4.71	1203
35	2.121	221	5.06	1216
36	2.400	224	5.54	1027
37	2.745	227	6.26	828
38	3.209	232	7.48	568
39	3.963	240	10.30	308
40	5.210	252	18.43	124

## **16. Validity**

### **16.1 External Third-Party Content Alignment Study**

The external third-party content alignment study was conducted in the State of Idaho for grades 3–8 and 10 in reading and mathematics and grades 5, 7, and 10 in science. The alignment study was held in Boise, Idaho, May 22–25, 2007. The breakdown of reviewers can be found in Appendix Y. The final report of study results consisted of a description of the four criteria used to judge the alignment between Idaho content objectives and the test questions of the reading, mathematics, and science ISAT. Each report included tables listing the results from the review process.

#### **Reviewers**

Each content alignment group consisted of national content experts and content experts from the State of Idaho. The State of Idaho reviewers have extensive teaching experience in the state and expertise in the field of reading, mathematics, or science, while the national reviewers have extensive expertise in the fields of reading, mathematics, or science standards, curriculum, and/or assessment design. Participants analyzed the alignment between the state's ISAT content standards and objectives and the test questions on the ISAT. The reviewers determined the alignment between the test questions and the five content standards in both mathematics and science and the two content standards in reading. A national psychometrician reviewed the final results of the alignment study. The results indicated that there is alignment between the content standards, goals, and objectives for reading and mathematics grades 3 through 8 and 10 and science grades 5, 7, and 10 and the test questions.

### **16.2 Content Validity**

#### **Content/Curricular**

The ISAT is composed of criterion-referenced assessments based on an extensive definition of the content assessed. Because it is content-based and aligned directly to the Idaho statewide content standards, it should demonstrate good content validity. Content validity addresses whether the test adequately samples the relevant material it purports to cover.

#### **Relation to Statewide Content Standards**

From the inception of the ISAT, committees of educators, item development experts, assessment experts, and OSBE staff have met periodically to review new and field tested items. A sequential review process has been put in place by OSBE. This provides many opportunities for these professionals to offer suggestions for improving or eliminating items as well as offer insights into the interpretation of the statewide content standards for the ISAT. These review committees participate in this process to ensure test content validity of the ISAT.

In addition to providing information on the difficulty, appropriateness, and sensitivity of these items, committee members provide a needed check on the alignment between the items and the content standards they are intended to measure. When items are judged relevant, that is, representative of the content defined by the standards, this judgment provides evidence to support the validity of inferences made (regarding knowledge of this content) with ISAT results. When items are judged to be inappropriate for any reason, the committee can either suggest revisions (e.g., reclassification, rewording) or elect to

eliminate the item from the field test item pool. Items that are approved by the review committee are later embedded in the operational ISAT form to allow for the collection of performance data. In essence, these committees review and verify the alignment of the test items with the objectives and measurement specifications to ensure that the items measure appropriate content. The nature and specificity of these review procedures provide strong evidence for the content validity of the ISAT.

## **Educator Input**

Idaho educators provide valuable input on the alignment of the items and the statewide content standards. Items are written specifically to measure the objectives and specifications of the content standards for the ISAT. Using a varied source of item writers provides a system of checks-and-balances for item development and review that reduces single source bias. Because many different people with different backgrounds write the items, it is less likely that items will suffer from a bias that might occur if items were written by a single author. This direct input from educators offers evidence regarding the content validity of the ISAT.

## **Developer Input**

For the items included in the 2007 form, OSBE and DRC staff provided a history of test building experience, including content-related expertise. The input and review by these assessment professionals provided further support of the item being an accurate measure of the intended objective. Thus, these reviews offer additional evidence for the content-validity of the ISAT.

## **Item to Content Area Match**

Expert judgments from educators, test developers, and assessment specialists provide support for the alignment of the ISAT with the statewide content standards. In addition, because expert teachers in the content areas were involved in establishing the content standards, the judgments of these same expert teachers in the review process provide a measure of content validity. A match between the content standards and the components of the ISAT provides evidence that the assessment measures the content standards. A table showing the number of assessment components, tasks, or items matching each content-standard is often used to provide documentation of the content validity of an assessment. The ISAT test blueprint provides this documentation. The blueprints for reading, mathematics, and writing in grades 2–10 are presented in Appendix V.

## **16.3 Intercorrelations-Convergent/Divergent Validity**

Tables 16.1 to 16.8 show the scale score intercorrelations for reading, mathematics, language usage, and science (grades 5, 7, and 10) for each grade. Disattenuated correlations (Spearman, 1904) which account for the unreliability of the measures, are also presented. A consistent pattern in the data is evident: reading and mathematics correlate the lowest of the pairs and reading and language usage correlate the highest. The latter pair shows an even higher disattenuated correlation primarily because of the somewhat lower reliability of language usage. Science pairs tend to fall in the middle.

**Table 16.1. Grade 3: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 3 Scale Score Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.682	0.725
Mathematics	0.682	1.000	0.707
Language Usage	0.725	0.707	1.000

<b>Grade 3 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.765	0.826
Mathematics	0.765	1.000	0.807
Language Usage	0.826	0.807	1.000

**Table 16.2. Grade 4: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 4 Scale Score Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.670	0.728
Mathematics	0.670	1.000	0.701
Language Usage	0.728	0.701	1.000

<b>Grade 4 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.764	0.869
Mathematics	0.764	1.000	0.842
Language Usage	0.869	0.842	1.000

**Table 16.3. Grade 5: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 5 Scale Score Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.693	0.757	0.721
Mathematics	0.693	1.000	0.735	0.666
Language Usage	0.757	0.735	1.000	0.680
Science	0.721	0.666	0.680	1.000

<b>Grade 5 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.770	0.870	0.842
Mathematics	0.770	1.000	0.844	0.778
Language Usage	0.870	0.844	1.000	0.821
Science	0.842	0.778	0.821	1.000

**Table 16.4. Grade 6: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 6 Scale Score Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.711	0.764
Mathematics	0.711	1.000	0.734
Language Usage	0.764	0.734	1.000

<b>Grade 6 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.802	0.882
Mathematics	0.802	1.000	0.839
Language Usage	0.882	0.839	1.000

**Table 16.5. Grade 7: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 7 Scale Score Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.727	0.784	0.769
Mathematics	0.727	1.000	0.739	0.743
Language Usage	0.784	0.739	1.000	0.731
Science	0.769	0.743	0.731	1.000

<b>Grade 7 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.804	0.896	0.866
Mathematics	0.804	1.000	0.837	0.829
Language Usage	0.896	0.837	1.000	0.843
Science	0.866	0.829	0.843	1.000

**Table 16.6. Grade 8: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 8 Scale Score Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.691	0.767
Mathematics	0.691	1.000	0.713
Language Usage	0.767	0.713	1.000

<b>Grade 8 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.774	0.884
Mathematics	0.774	1.000	0.814
Language Usage	0.884	0.814	1.000



**Table 16.7. Grade 9: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 9 Scale Score Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.667	0.704
Mathematics	0.667	1.000	0.728
Language Usage	0.704	0.728	1.000

<b>Grade 9 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage
Reading	1.000	0.766	0.852
Mathematics	0.766	1.000	0.849
Language Usage	0.852	0.849	1.000

**Table 16.8. Grade 10: Scale Score Correlation and Disattenuated Correlation**

<b>Grade 10 Scale Score Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.699	0.755	0.749
Mathematics	0.699	1.000	0.701	0.723
Language Usage	0.755	0.701	1.000	0.707
Science	0.749	0.723	0.707	1.000

<b>Grade 10 Scale Score Disattenuated Correlation</b>	Reading	Mathematics	Language Usage	Science
Reading	1.000	0.783	0.878	0.835
Mathematics	0.783	1.000	0.817	0.808
Language Usage	0.878	0.817	1.000	0.820
Science	0.835	0.808	0.820	1.000

Tables 16.9 to 16.12 show the raw score intercorrelations for the content standard subscores. While the subscore correlation is shown for each subscore with the total raw score, given the overlap, these correlations are spuriously inflated.

For reading, the correlations all fall in the mid-70's, indicating that the two subscores share approximately half of their variance. This indicates some evidence that the two subscores are not measuring the same thing. That is, there is a fair amount of unique variance.

For mathematics, the correlations are lower than those for reading, falling into a range from .317 to .742. The correlations are the lowest for Concepts and Principles of Measurement with Concepts and Principles of Geometry in grade 3 through 8 and highest for Number and Operation with Concepts and Language of Algebra and Functions for all grades. In general, the intercorrelations tend to be higher in the upper grades than in the lower grades.

For language usage, the correlations between writing process and writing components ranges from .635 in grade 4 to .714 in grade 3.

For science in grades 5 and 7, Nature of Science subscores correlate higher than the other pair-wise correlations (Physical Science, Biology, Earth and Space Systems, and Personal and Social perspectives; Technology). In grade 10, Nature of Science as well as Personal and Social Perspectives; Technology correlated higher than the remaining subscore pairs.

All of the correlations suggest that each subscore contributes significantly and uniquely to its respective total score.

**Table 16.9. Reading Intercorrelations**

<b>Grade 3 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.900	0.962
Reading Process	0.900	1.000	0.746
Comprehension/Interpretation	0.962	0.746	1.000

<b>Grade 4 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.865	0.975
Reading Process	0.865	1.000	0.734
Comprehension/Interpretation	0.975	0.734	1.000

<b>Grade 5 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.873	0.981
Reading Process	0.873	1.000	0.763
Comprehension/Interpretation	0.981	0.763	1.000

<b>Grade 6 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.866	0.975
Reading Process	0.866	1.000	0.735
Comprehension/Interpretation	0.975	0.735	1.000

**Table 16.9 (continued). Reading Intercorrelations**

<b>Grade 7 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.852	0.981
Reading Process	0.852	1.000	0.735
Comprehension/Interpretation	0.981	0.735	1.000

<b>Grade 8 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.857	0.980
Reading Process	0.857	1.000	0.737
Comprehension/Interpretation	0.980	0.737	1.000

<b>Grade 9 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.812	0.984
Reading Process	0.812	1.000	0.693
Comprehension/Interpretation	0.984	0.693	1.000

<b>Grade 10 Reading</b>	Total	Reading Process	Comprehension/Interpretation
Total	1.000	0.864	0.984
Reading Process	0.864	1.000	0.762
Comprehension/Interpretation	0.984	0.762	1.000

**Table 16.10. Mathematics Intercorrelations**

<b>Grade 3 Mathematics</b>	Total	Number and Operation	Concepts and Principles of Measurement	Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.909	0.751	0.782	0.721	0.781
Number and Operation	0.909	1.000	0.578	0.639	0.576	0.612
Concepts and Principles of Measurement	0.751	0.578	1.000	0.503	0.458	0.491
Concepts and Language of Algebra and Functions	0.782	0.639	0.503	1.000	0.486	0.528
Concepts and Principles of Geometry	0.721	0.576	0.458	0.486	1.000	0.518
Data Analysis, Probability, and Statistics	0.781	0.612	0.491	0.528	0.518	1.000

<b>Grade 4 Mathematics</b>	Total	Number and Operation	Concepts and Principles of Measurement	Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.899	0.725	0.770	0.592	0.768
Number and Operation	0.899	1.000	0.559	0.614	0.386	0.600
Concepts and Principles of Measurement	0.725	0.559	1.000	0.484	0.317	0.483
Concepts and Language of Algebra and Functions	0.770	0.614	0.484	1.000	0.333	0.520
Concepts and Principles of Geometry	0.592	0.386	0.317	0.333	1.000	0.371
Data Analysis, Probability, and Statistics	0.768	0.600	0.483	0.520	0.371	1.000

**Table 16.10 (continued). Mathematics Intercorrelations**

<b>Grade 5 Mathematics</b>	Total	Number and Operation	Concepts and Principles of Measurement	Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.919	0.808	0.791	0.767	0.785
Number and Operation	0.919	1.000	0.693	0.651	0.610	0.648
Concepts and Principles of Measurement	0.808	0.693	1.000	0.566	0.530	0.545
Concepts and Language of Algebra and Functions	0.791	0.651	0.566	1.000	0.512	0.549
Concepts and Principles of Geometry	0.767	0.610	0.530	0.512	1.000	0.543
Data Analysis, Probability, and Statistics	0.785	0.648	0.545	0.549	0.543	1.000

<b>Grade 6 Mathematics</b>	Total	Number and Operation	Concepts and Principles of Measurement	Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.886	0.795	0.858	0.782	0.783
Number and Operation	0.886	1.000	0.641	0.693	0.595	0.616
Concepts and Principles of Measurement	0.795	0.641	1.000	0.623	0.526	0.555
Concepts and Language of Algebra and Functions	0.858	0.693	0.623	1.000	0.570	0.600
Concepts and Principles of Geometry	0.782	0.595	0.526	0.570	1.000	0.538
Data Analysis, Probability, and Statistics	0.783	0.616	0.555	0.600	0.538	1.000

**Table 16.10 (continued). Mathematics Intercorrelations**

<b>Grade 7 Mathematics</b>				Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.913	0.817	0.852	0.789	0.802
Number and Operation	0.913	1.000	0.690	0.720	0.619	0.680
Concepts and Principles of Measurement	0.817	0.690	1.000	0.625	0.571	0.590
Concepts and Language of Algebra and Functions	0.852	0.720	0.625	1.000	0.590	0.616
Concepts and Principles of Geometry	0.789	0.619	0.571	0.590	1.000	0.543
Data Analysis, Probability, and Statistics	0.802	0.680	0.590	0.616	0.543	1.000

<b>Grade 8 Mathematics</b>				Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.886	0.689	0.907	0.782	0.775
Number and Operation	0.886	1.000	0.532	0.742	0.611	0.633
Concepts and Principles of Measurement	0.689	0.532	1.000	0.533	0.454	0.460
Concepts and Language of Algebra and Functions	0.907	0.742	0.533	1.000	0.616	0.642
Concepts and Principles of Geometry	0.782	0.611	0.454	0.616	1.000	0.524
Data Analysis, Probability, and Statistics	0.775	0.633	0.460	0.642	0.524	1.000

**Table 16.10 (continued). Mathematics Intercorrelations**

<b>Grade 9 Mathematics</b>				Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.826	0.861	0.903	0.771	0.785
Number and Operation	0.826	1.000	0.624	0.698	0.551	0.578
Concepts and Principles of Measurement	0.861	0.624	1.000	0.693	0.606	0.622
Concepts and Language of Algebra and Functions	0.903	0.698	0.693	1.000	0.615	0.632
Concepts and Principles of Geometry	0.771	0.551	0.606	0.615	1.000	0.524
Data Analysis, Probability, and Statistics	0.785	0.578	0.622	0.632	0.524	1.000

<b>Grade 10 Mathematics</b>				Concepts and Language of Algebra and Functions	Concepts and Principles of Geometry	Data Analysis, Probability, and Statistics
Total	1.000	0.819	0.785	0.901	0.840	0.736
Number and Operation	0.819	1.000	0.578	0.675	0.614	0.513
Concepts and Principles of Measurement	0.785	0.578	1.000	0.629	0.603	0.489
Concepts and Language of Algebra and Functions	0.901	0.675	0.629	1.000	0.670	0.566
Concepts and Principles of Geometry	0.840	0.614	0.603	0.670	1.000	0.553
Data Analysis, Probability, and Statistics	0.736	0.513	0.489	0.566	0.553	1.000

**Table 16.11. Language Usage Intercorrelations**

<b>Grade 3 Language Usage</b>			
	Total	Writing Process	Writing Components
Total	1.000	0.897	0.950
Writing Process	0.897	1.000	0.714
Writing Components	0.950	0.714	1.000

<b>Grade 4 Language Usage</b>			
	Total	Writing Process	Writing Components
Total	1.000	0.874	0.930
Writing Process	0.874	1.000	0.635
Writing Components	0.930	0.635	1.000

<b>Grade 5 Language Usage</b>			
	Total	Writing Process	Writing Components
Total	1.000	0.927	0.914
Writing Process	0.927	1.000	0.694
Writing Components	0.914	0.694	1.000

<b>Grade 6 Language Usage</b>			
	Total	Writing Process	Writing Components
Total	1.000	0.937	0.906
Writing Process	0.937	1.000	0.701
Writing Components	0.906	0.701	1.000

<b>Grade 7 Language Usage</b>			
	Total	Writing Process	Writing Components
Total	1.000	0.948	0.884
Writing Process	0.948	1.000	0.690
Writing Components	0.884	0.690	1.000



**Table 16.11 (continued). Language Usage Intercorrelations**

<b>Grade 8 Language Usage</b>	Total	Writing Process	Writing Components
Total	1.000	0.951	0.877
Writing Process	0.951	1.000	0.687
Writing Components	0.877	0.687	1.000

<b>Grade 9 Language Usage</b>	Total	Writing Process	Writing Components
Total	1.000	0.947	0.861
Writing Process	0.947	1.000	0.651
Writing Components	0.861	0.651	1.000

<b>Grade 10 Language Usage</b>	Total	Writing Process	Writing Components
Total	1.000	0.945	0.859
Writing Process	0.945	1.000	0.645
Writing Components	0.859	0.645	1.000

**Table 16.12. Science Intercorrelations**

<b>Grade 5 Science</b>					Earth and Space Systems	Personal and Social Perspectives; Technology
Total	1.000	0.875	0.707	0.687	0.593	0.736
Nature of Science	0.875	1.000	0.509	0.463	0.390	0.571
Physical Science	0.707	0.509	1.000	0.401	0.327	0.422
Biology	0.687	0.463	0.401	1.000	0.304	0.385
Earth and Space Systems	0.593	0.390	0.327	0.304	1.000	0.304
Personal and Social Perspectives; Technology	0.736	0.571	0.422	0.385	0.304	1.000

<b>Grade 7 Science</b>					Earth and Space Systems	Personal and Social Perspectives; Technology
Total	1.000	0.935	0.738	0.768	0.711	0.697
Nature of Science	0.935	1.000	0.608	0.640	0.581	0.562
Physical Science	0.738	0.608	1.000	0.491	0.450	0.425
Biology	0.768	0.640	0.491	1.000	0.468	0.439
Earth and Space Systems	0.711	0.581	0.450	0.468	1.000	0.418
Personal and Social Perspectives; Technology	0.697	0.562	0.425	0.439	0.418	1.000

**Table 16.12 (continued). Science Intercorrelations**

<b>Grade 10 Science</b>		Nature of Science	Physical Science	Biology	Earth and Space Systems	Personal and Social Perspectives; Technology
Total	1.000	0.942	0.671	0.756	0.782	0.855
Nature of Science	0.942	1.000	0.546	0.634	0.663	0.746
Physical Science	0.671	0.546	1.000	0.429	0.435	0.481
Biology	0.756	0.634	0.429	1.000	0.535	0.591
Earth and Space Systems	0.782	0.663	0.435	0.535	1.000	0.624
Personal and Social Perspectives; Technology	0.855	0.746	0.481	0.591	0.624	1.000

## ***16.4 Relationship Between the Core and Extender Scores***

### **Descriptives**

For each grade and content combination, Table 16.13 shows the average scale scores of the core and core-plus-extender and the correlation between the two. Given that the items in the core made up from 71.4 percent (grade 3 reading) to 78.1 percent (grade 10 science) of the core-plus-extender item set, the correlations are expected to be high. As for the differences between the means of the two scores, these should be randomly distributed around zero.

**Table 16.13. Core and Extender Scale Scores Summary**

<b>Content</b>	<b>Grade</b>	<b>Average Scale Score Core</b>	<b>Average Scale Score Extender</b>	<b>Correlation</b>
Reading	3	203.0	203.1	0.966
Reading	4	207.8	207.8	0.971
Reading	5	213.3	212.9	0.972
Reading	6	216.1	216.2	0.970
Reading	7	220.7	220.4	0.972
Reading	8	225.3	224.7	0.972
Reading	10	227.5	226.8	0.969
Mathematics	3	228.6	228.2	0.976
Mathematics	4	204.3	205.0	0.944
Mathematics	5	211.1	211.4	0.959
Mathematics	6	218.4	218.2	0.973
Mathematics	7	225.8	225.4	0.973
Mathematics	8	230.5	230.3	0.977
Mathematics	10	236.3	236.0	0.978
Language Usage	3	242.6	241.3	0.983
Language Usage	4	244.3	243.7	0.977
Language Usage	5	201.2	202.1	0.950
Language Usage	6	210.1	209.2	0.961
Language Usage	7	213.3	213.0	0.970
Language Usage	8	218.4	217.5	0.969
Language Usage	10	220.8	220.1	0.975
Science	5	223.6	222.7	0.971
Science	7	225.2	224.6	0.975
Science	10	228.4	227.5	0.976

## Decision Consistency

Another way to look at the impact of the extender section on a criterion-referenced test is to evaluate the consistency of the placement of students into the four proficiency levels. Note that for student and AYP reporting purposes, the core set alone was used. In this sense, the “what if” is not whether an alternate form might produce a different result as is the case with many of these types of comparisons. Rather, it is whether a longer test with more of the same items might produce a different classification. For all intents and purposes, this comparison has little application in terms of the assessment. As stated in a previous section, the reliability of the ISAT is sufficiently high without the extender. However, the extent to which the decision consistency is high lends some degree of validity to the extender in terms of it measuring the same construct as the core.

Tables 16.14 to 16.40 show four-by-four contingency tables, with each cell representing the cross between the core proficiency level and the core-plus-extender proficiency level. The diagonals represent

agreement and the adjacent-to-diagonals represent adjacent agreement (within one level). Any non-zero cells outside of these represent changes in the categorization under the two models (core versus core-plus-extender) of two levels of more. The percent perfect agreement is presented at the bottom of each table, along with Cohen's Kappa (Cohen, 1960). Kappa may be interpreted as the proportion of agreement after chance agreement has been excluded.

Clearly, the exact agreement is quite high, ranging from 87.0 to 89.4 for reading, 86.7 to 89.4 for mathematics, 79.6 (grade 3) to 89.6 for language usage, and 86.3 to 88.4 in science. The kappas, ranging from .713 to .846, also show a strong association.

Upon further investigation of those students whose proficiency level changed by two levels, it was found that their performance varied by a significant amount between the core section and extender section. For example, one student scored 13 out of 40 on the core section and 12 out of 16 on the extender. In this case, they were classified as *just barely* Below Basic on the core and *just barely* Proficient on the combined core and extender (a raw score of 25 out of 56). This pattern was evident in other, similar cases.

**Table 16.14. Grade 3 Reading – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	1392	393	2	0	1787
	Basic	109	1375	511	0	1995
	Proficient	0	120	7749	696	8565
	Advanced	0	0	743	6659	7402
	Total	1501	1888	9005	7355	

Agreement = 87.0%

Cohen Kappa = 0.798

**Table 16.15. Grade 4 Reading – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	1406	322	3	0	1731
	Basic	273	1361	405	0	2039
	Proficient	5	374	8873	613	9865
	Advanced	0	0	495	5372	5867
	Total	1684	2057	9776	5985	

Agreement = 87.2%

Cohen Kappa = 0.799

**Table 16.16. Grade 5 Reading – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1479	400	0	0	1879
	<b>Basic</b>	146	1948	266	0	2360
	<b>Proficient</b>	0	347	8312	636	9295
	<b>Advanced</b>	0	0	506	5729	6235
	<b>Total</b>	1625	2695	9084	6365	

Agreement = 88.4%

Cohen Kappa = 0.823

**Table 16.17. Grade 6 Reading – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1282	357	0	0	1639
	<b>Basic</b>	173	2105	461	0	2739
	<b>Proficient</b>	1	259	8830	681	9771
	<b>Advanced</b>	0	0	493	4751	5244
	<b>Total</b>	1456	2721	9784	5432	

Agreement = 87.5%

Cohen Kappa = 0.806

**Table 16.18. Grade 7 Reading – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1284	198	0	0	1482
	<b>Basic</b>	243	2416	357	0	3016
	<b>Proficient</b>	0	480	7971	411	8862
	<b>Advanced</b>	0	0	646	5648	6294
	<b>Total</b>	1527	3094	8974	6059	

Agreement = 88.1%

Cohen Kappa = 0.821

**Table 16.19. Grade 8 Reading – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	597	96	0	0	693
	<b>Basic</b>	313	1592	209	0	2114
	<b>Proficient</b>	1	419	8700	362	9482
	<b>Advanced</b>	0	0	703	6808	7511
	<b>Total</b>	911	2107	9612	7170	

Agreement = 89.4%

Cohen Kappa = 0.828

**Table 16.20. Grade 9 Reading – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	517	88	0	0	605
	Basic	372	1444	83	0	1899
	Proficient	0	732	9757	600	11089
	Advanced	0	0	685	6027	6712
	Total	889	2264	10525	6627	

Agreement = 87.4%

Cohen Kappa = 0.789

**Table 16.21. Grade 10 Reading – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	984	244	0	0	1228
	Basic	186	2325	216	0	2727
	Proficient	0	487	8021	396	8904
	Advanced	0	0	478	5343	5821
	Total	1170	3056	8715	5739	

Agreement = 89.3%

Cohen Kappa = 0.836

**Table 16.22. Grade 3 Mathematics – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	289	183	0	0	472
	Basic	40	1816	374	0	2230
	Proficient	0	140	5820	1417	7377
	Advanced	0	0	225	9572	9797
	Total	329	2139	6419	10989	

Agreement = 88.0%

Cohen Kappa = 0.799

**Table 16.23. Grade 4 Mathematics – Core Scale Score vs. Extender Scale Score**

		Extender				Total
		Below Basic	Basic	Proficient	Advanced	
Core	Below Basic	874	66	0	0	940
	Basic	334	2076	198	0	2608
	Proficient	0	441	7984	1140	9565
	Advanced	0	0	432	6107	6539
	Total	1208	2583	8614	7247	

Agreement = 86.7%

Cohen Kappa = 0.794

**Table 16.24. Grade 5 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1296	135	0	0	1431
	<b>Basic</b>	326	3345	257	0	3928
	<b>Proficient</b>	0	424	7141	578	8143
	<b>Advanced</b>	0	0	383	5976	6359
	<b>Total</b>	1622	3904	7781	6554	

Agreement = 89.4%

Cohen Kappa = 0.846

**Table 16.25. Grade 6 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1080	110	0	0	1190
	<b>Basic</b>	399	3195	122	0	3716
	<b>Proficient</b>	0	688	7233	587	8508
	<b>Advanced</b>	0	0	343	5702	6045
	<b>Total</b>	1479	3993	7698	6289	

Agreement = 88.4%

Cohen Kappa = 0.831

**Table 16.26. Grade 7 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1884	209	0	0	2093
	<b>Basic</b>	442	2995	357	0	3794
	<b>Proficient</b>	0	348	7062	691	8101
	<b>Advanced</b>	0	0	206	5498	5704
	<b>Total</b>	2326	3552	7625	6189	

Agreement = 88.6%

Cohen Kappa = 0.837

**Table 16.27. Grade 8 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1424	135	0	0	1559
	<b>Basic</b>	561	3236	243	0	4040
	<b>Proficient</b>	0	648	7242	552	8442
	<b>Advanced</b>	0	0	348	5439	5787
	<b>Total</b>	1985	4019	7833	5991	

Agreement = 87.5%

Cohen Kappa = 0.819



**Table 16.28. Grade 9 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1634	69	0	0	1703
	<b>Basic</b>	803	1992	119	0	2914
	<b>Proficient</b>	9	867	7489	184	8549
	<b>Advanced</b>	0	0	626	6747	7373
	<b>Total</b>	2446	2928	8234	6931	

Agreement = 87.0%

Cohen Kappa = 0.809

**Table 16.29. Grade 10 Mathematics – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1795	114	0	0	1909
	<b>Basic</b>	513	2581	140	0	3234
	<b>Proficient</b>	2	908	7055	401	8366
	<b>Advanced</b>	0	0	420	5010	5430
	<b>Total</b>	2310	3603	7615	5411	

Agreement = 86.8%

Cohen Kappa = 0.810

**Table 16.30. Grade 3 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	2154	375	2	0	2531
	<b>Basic</b>	138	2659	1296	0	4093
	<b>Proficient</b>	0	107	5181	1619	6907
	<b>Advanced</b>	0	0	510	5797	6307
	<b>Total</b>	2292	3141	6989	7416	

Agreement = 79.6%

Cohen Kappa = 0.713

**Table 16.31. Grade 4 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	972	81	0	0	1053
	<b>Basic</b>	543	2112	204	0	2859
	<b>Proficient</b>	3	894	8522	393	9812
	<b>Advanced</b>	0	0	1110	4763	5873
	<b>Total</b>	1518	3087	9836	5156	

Agreement = 83.5%

Cohen Kappa = 0.744

**Table 16.32. Grade 5 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1657	143	0	0	1800
	<b>Basic</b>	438	3445	491	0	4374
	<b>Proficient</b>	0	376	8949	673	9998
	<b>Advanced</b>	0	0	473	3197	3670
	<b>Total</b>	2095	3964	9913	3870	

Agreement = 86.9%

Cohen Kappa = 0.801

**Table 16.33. Grade 6 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1943	118	0	0	2061
	<b>Basic</b>	482	3402	373	0	4257
	<b>Proficient</b>	0	540	8416	448	9404
	<b>Advanced</b>	0	0	619	3074	3693
	<b>Total</b>	2425	4060	9408	3522	

Agreement = 86.7%

Cohen Kappa = 0.802

**Table 16.34. Grade 7 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	1905	108	0	0	2013
	<b>Basic</b>	468	4193	209	0	4870
	<b>Proficient</b>	0	773	9325	382	10480
	<b>Advanced</b>	0	0	379	1910	2289
	<b>Total</b>	2373	5074	9913	2292	

Agreement = 88.2%

Cohen Kappa = 0.816

**Table 16.35. Grade 8 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	2343	76	0	0	2419
	<b>Basic</b>	554	4183	411	0	5148
	<b>Proficient</b>	0	544	9200	92	9836
	<b>Advanced</b>	0	0	1030	1361	2391
	<b>Total</b>	2897	4803	10641	1453	

Agreement = 86.3%

Cohen Kappa = 0.787

**Table 16.36. Grade 9 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	2309	153	0	0	2462
	<b>Basic</b>	724	4934	254	0	5912
	<b>Proficient</b>	0	926	9832	228	10986
	<b>Advanced</b>	0	0	157	925	1082
	<b>Total</b>	3033	6013	10243	1153	

Agreement = 88.1%

Cohen Kappa = 0.809

**Table 16.37. Grade 10 Language Usage – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	2055	141	0	0	2196
	<b>Basic</b>	467	3891	224	0	4582
	<b>Proficient</b>	1	749	10324	91	11165
	<b>Advanced</b>	0	0	302	743	1045
	<b>Total</b>	2523	4781	10850	834	

Agreement = 89.6%

Cohen Kappa = 0.822

**Table 16.38. Grade 5 Science – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	806	394	0	0	1200
	<b>Basic</b>	72	7218	705	0	7995
	<b>Proficient</b>	0	640	6837	337	7814
	<b>Advanced</b>	0	0	560	2206	2766
	<b>Total</b>	878	8252	8102	2543	

Agreement = 86.3%

Cohen Kappa = 0.789

**Table 16.39. Grade 7 Science – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
<b>Core</b>	<b>Below Basic</b>	4971	375	0	0	5346
	<b>Basic</b>	438	3933	514	0	4885
	<b>Proficient</b>	0	433	3477	606	4516
	<b>Advanced</b>	0	0	270	4545	4815
	<b>Total</b>	5409	4741	4261	5151	

Agreement = 86.5%

Cohen Kappa = 0.820

**Table 16.40. Grade 10 Science – Core Scale Score vs. Extender Scale Score**

		<b>Extender</b>				<b>Total</b>
<b>Core</b>		<b>Below Basic</b>	<b>Basic</b>	<b>Proficient</b>	<b>Advanced</b>	
	<b>Below Basic</b>	4286	256	0	0	4542
	<b>Basic</b>	443	2564	184	0	3191
	<b>Proficient</b>	0	622	6371	378	7371
	<b>Advanced</b>	0	0	265	3166	3431
	<b>Total</b>	4729	3442	6820	3544	

Agreement = 88.4%

Cohen Kappa = 0.840

## **17. Proficiency Level Descriptors and Standards Validation**

### **17.1 Proficiency Level Descriptors**

The proficiency level descriptors for reading and mathematics were developed by a group of Idaho educators facilitated by DRC content area specialists in July 2006. Committee members were provided initial training that included background information about the Idaho content standards, goals, and objectives; the purpose and use of the proficiency level descriptors; and the process that would be used to develop the descriptors. Following the training, the committee members were divided into subcommittees by content area for reading and mathematics. Each content committee consisted of 35 Idaho educators and one DRC content specialist. The content committees first worked together on the grade 6 proficiency level descriptors and were then divided into two groups – fifth grade and below and seventh grade and above. The subcommittees reviewed and revised draft preliminary proficiency level descriptors that had been prepared by DRC's content leads using the Idaho content standards, goals, and objectives. The draft preliminary proficiency level descriptors were revised to:

- clearly define what students in Idaho should know and be able to do based on the goals and objectives,
- clearly show how students would demonstrate this knowledge and skill based on the goal and objectives, and
- clearly define the level of knowledge and skill necessary for each proficiency level.

Afterwards, the content subcommittee reconvened as a whole group to review the recommended changes of each subcommittee, and the committees approved the final proficiency level descriptors.

The proficiency level descriptors for reading and mathematics were revisited by a committee of Idaho educators in March 2007. Again the committee of Idaho educators reviewed and revised the preliminary proficiency level descriptors. In addition, language usage and science draft preliminary proficiency level descriptors were prepared by DRC's content leads and were reviewed and revised at the meeting. Group training was provided that included background information about the content standards, goals, and objectives; the purpose and use of the proficiency level descriptors; and the process that was used to develop the descriptors. After the training, the committee members were divided into subcommittees by content area of reading, mathematics, language usage, and science. The subcommittees reviewed and revised the draft preliminary proficiency level descriptors.

### **17.2 Standards Validation**

See Appendix W for the *Standards Validation Report*.

## References

- Allen, N. L., Carlson J. E., & Zelenak, C. A., (1999). The NAEP 1996 technical report (NCES 1999–452). Washington, DC: National Center for Education Statistics.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing (2nd ed.). Washington, DC: American Educational Research Association.
- Cohen, J. (1960). “A coefficient of agreement for nominal scales.” Educational and Psychological Measurement. 20: 37-46.
- Cronbach, L. (1951). Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16, 297-334.
- Huynh, H. (2000). Technical Documentation for the Palmetto Achievement Challenge Test.
- Huynh, H and Schneider, C (2004). Vertically Moderated Standards as an Alternative to Vertical Scaling: Assumptions, Practices, and an Odyssey through NAEP. Paper presented at the annual meeting of the National Council on Measurement in Education, April 2004, San Diego, CA.
- Linacre, John M. 2002. A User’s Guide to WINSTEPS. Chicago: Mesa Press.
- Linacre, J. M. 2006. WINSTEPS Rasch measurement (Version 3.63). Chicago: WINSTEPS.com. Computer program.
- Livingston, S. & Lewis, C. (1995). “Estimating the Consistency and Accuracy of Classifications Based on Test Scores.” Journal of Educational Measurement 32, 179-197.
- Mantel, Nathan, and Haenszel, William. 1959. “Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease.” Journal of the National Cancer Institute 22:719–48.
- Rasch, Georg. 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danish Institute for Educational Research.
- Spearman, C (1904). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.
- Thompson, S., Johnston, C. J., and Thurlow, M. L. (2002). Universal design applied to large scale assessments. National Center on Educational Outcomes Synthesis Report 44. Minneapolis, MN: University of Minnesota.



